

Can Fortress Europe be feasible?

The governments of Germany and France have, in different ways, raised questions about the future of European industry that deserve better answers than they are likely to be given.

JUST over half a year ago, there was general celebration at the reunification of what were then East and West Germany. Nobody then imagined that last week would see the assassination of Detlev Rohwedder, the man who did the thankless job of administering the *Treuhand* — a kind of holding company for the state-run industrial and commercial enterprises of the east which he converted into a disposal agency. The assassination is personally tragic, of course: why should a public servant (and his family and friends) be required to pay with his life for the difficulty of the job he undertakes? It is also tragic that violence of a kind that has been abroad in West Germany for 30 years should now have appeared in the east. But it is just as alarming that bright hopes of making the eastern parts of Germany prosperous by industrial development have so quickly foundered.

The explanation, as the chairman of the Bundesbank seems not to tire of saying, is elementary. When Bonn decided that Deutschmarks would instantly become the accepted currency in old East Germany, and when wages there had been pegged to rates in the West (without which emigration would have continued and even accelerated), it was inevitable that most eastern enterprises would be unable to compete with their western counterparts. So it has proved. State-owned companies are collapsing before the *Treuhand* can get them off its hands. Unemployment is rife, and there are crowds marching in city streets. Some guess that half the eastern workforce may be unemployed later in the year. Meanwhile, taxes in Germany have been increased to pay for the subventions and the capital investment necessary to keep the east afloat. Five years from now, there may be a different tale to tell, but five years is a long time for people on the dole.

The developments in France are entirely different and unrelated, but provoke a common question. France, the self-nominated guardian of European technology (with its public interests in aerospace and electronics) is alarmed that two important companies, the computer manufacturer Bull and the electronics and defence manufacturer Thomson, have run into financial trouble, and is planning to help them out with substantial amounts of capital on a scale likely to offend against European rules on unfair competition. France makes no secret of its ambitions: to safeguard in Europe the essential parts of modern industry against the threat of competition from the Far East (not exclusively Japan) and from North America. What the government of France should ask itself is why it is better for the European cause that it should invest in Bull and Thomson than in, say, eastern Germany.

In Europe, France is about as firmly committed as any other member of the European Communities to monetary and political union along lines now being argued out in the committees of the two intergovernmental conferences on these subjects. France must assume that, by about the turn of the century, Europe will be a true common market and even if it does not have a common currency, will have developed a system of national and social mechanisms to ensure that European currencies retain their relative value. And whatever happens, there is no prospect that any single member will think it sensible to get so entirely out of step politically with its fellow-members that these financial arrangements are put at risk. In short, if France were to invest in eastern Germany, it need not fear that at some later stage its investments would be expropriated.

But how would that help to keep advanced technology alive in Europe? This way. If a sixth of Germany is destined to be underemployed while waiting for investment to catch up with the present capital requirements of now-familiar jobs, Europe will be less rich than otherwise. So important social services (education, for example) are needlessly neglected, the European market languishes longer than is necessary and Europe as a whole earns less than it might in the export trade. So, five or ten years from now, Europe will be less able than would otherwise have been the case to invest in the then fashionable technology. Investment in eastern Germany might well, in other words, be at once more profitable and a better defence of European competence than the planned investment in Bull and Thomson — and also less troublesome for Brussels. Who will tell that to the French? □

Australia's talisman

The Canberra government seems to be playing fast and loose with the Australian National University.

Dom Mintoff, when prime minister of Malta in the 1970s, hit on a simple but effective way of cutting down to size the irritatingly distinguished University of Malta: he used the power of his government's purse to force its merger with the polytechnic institution in the city of Valletta. The predictable result was also quick; true academics quit, would-be academics eagerly filled their shoes, some in Malta cherish the memory of having taken an axe to academic privilege and Malta lacks an outstanding university. In Australia, Bob Hawke's gov-



ernment has wisely (if only just) shrunk from following Min-toff's recipe, but the effect of its proposals for research at the Australian National University (ANU), announced two weeks ago, may be much the same (page 452).

ANU is not merely privileged, but constitutionally so: the bulk of its research costs are met directly by the federal government, by-passing the usual grant-making procedures. The rationale derives from the conviction of past Australian governments of a need for a centre of some excellence and for means of training some graduate students indigenously.

Two things have changed decisively, and to ANU's disadvantage, in the past four decades — Australia has become even less respectful of entrenched privilege than its admirably disrespectful traditions would allow, while the rapid government-inspired growth of higher education has endowed with international standing between three and four other universities. (Subscription-conscious journals cannot be expected to be more precise about the number, or to say which they are.) ANU, which has seemed vulnerable for at least the past ten years, would now have been able to defend itself more strongly if the quality of its research had spilled more generously into its teaching.

The new proposals nevertheless promise needless damage. By requiring that each of the seven research schools at ANU should separately justify its continued existence, they will destroy ANU's ambition to be an all-round research centre of excellence. Worse, the separate schools will have to justify themselves on a playing-field made far from level by the decision that budgets should shrink in the interests (laudable though they are) of collaborative research. And can it be fair that the Australian Research Council, which stands to benefit from whatever research funds can be squirrelled away from ANU, should be the invigilator of the changes now decreed? Especially in Australia, that is asking a lot.

The underlying worry is that Australia's policy in higher education and research has become doctrinal, latterly determined by John S. Dawkins, the senior minister at the Department of Employment, Education and Training in Canberra. The objective, in what the prime minister calls 'the Clever Country', is admirable: higher education and advanced training for as many as possible. But uniform rules of thumb (on, for example, the economical size of a university) have taken precedence over a proper regard for the welfare of institutions, which can be most easily be crippled by being arbitrarily pushed about often enough.

Yet the diversity of support for higher education and research has also been shrinking ever since the early 1980s, when the federal government 'relieved' the states of the obligation to contribute to regional university finances. The constantly narrowing role of the Commonwealth Scientific Industrial Research Organization has made matters worse. Canberra bureaucrats, already too well-practised at second-guessing academics, may soon call all the shots that matter.

The future of ANU will (or should) be a test-case. If fears of what lies ahead are justified by events, as they may well be, ANU will cease to be an autonomous university, capable of deciding what it does corporately, in teaching and research. The complaint, by some, that it is not fulfilling its intended

role as a centre of excellence in research — both a magnet for overseas interest and an Australian national resource — will prove to be a self-justifying prophecy brought about by administrative means. But that is a roundabout way of reaching what should be an important national decision. □

Hell in a small place

Extinguishing the crop of burning oil wells in Kuwait may yet bring benefits to the oil-well business.

IN the war against the oil-well fires raging in Kuwait, there seem to be very few pacifists. The fires have captured the imagination not merely of the husky men from Houston, in the oil-patch, but also of many who have not seen an oil well on fire. Last week in Washington, the Union of Concerned Scientists, more used to attacking the military-industrial complex than joining in a war, used instant largesse from the Macarthur Foundation to organize a symposium on novel ways of putting out oil-well fires. Daedalus's DREADCO would have been proud to have been the sponsor. And, as if to emphasize that Kuwait's firefight is indeed a war, the symposium was preceded by a month-long bout of computer networking by militarily related research groups.

But the task in Kuwait will be complicated by the relics of the real war — the unexploded mines and bomblets richly strewn in some of the oil fields by the recent combatants. It seems to be agreed that even Texan hellfighters cannot be asked to put down their boots where there remains a real danger of their losing their feet. So there were schemes last week for using helicopters to drag mine-ploughs and harrows through the minefields; wistfully, some talked of leasing heavily armoured Soviet helicopters for this purpose.

Novel armchair schemes for putting out oil-well fires continue to multiply. Some last week were canvassing the idea of organizing implosions around cased holes in the ground and of using martensitic shape-memory alloy pipes to squash shut after being installed or to clamp onto fractured well-heads. Liquid nitrogen was much talked of, but schemes for bulk delivery to the desert left much to be desired. But the men from Houston kept bemoaning the lack of adequate water (now on the way to being resolved by the use of the pipeline network in reverse). Schemes for reducing environmental damage were also put forward last week. Making the wells burn free of soot would be an advantage; water injection might help to remove both soot and carcinogenic polycyclic hydrocarbons.

Meanwhile, it seems agreed that the 'mother of all fire experiments' is a poor approximation to a global doomsday machine: the 10 °C cooling beneath the plume argues for 'nuclear autumn' rather than 'winter'. And there must be some hollow laughter at the complaint of the Iraqi ambassador to the United Nations that depleted-uranium anti-tank shells used by US forces in the recent conflict may be an environmental hazard. The tonnage of raw uranium (not to mention vanadium) discharged from the burning oil wells far exceeds that left lying in the desert. □

NAS on global warming: 'insurance' needed

■ Report urges improved energy efficiency

■ Seeks research for adapting to greenhouse

Washington

Two-and-a-half years after Congress requested it, the US National Academy of Sciences (NAS) has finished its report on the *Policy Implications of Global Warming*. The Evans report (named for Daniel J. Evans, former senator from Washington State and chairman of the NAS panel) calls for a limited set of actions as 'insurance' against the possibility that an increasing amount of greenhouse gases in the atmosphere will raise the Earth's average temperature in coming decades.

The measures recommended by the NAS panel stick to the middle ground in the current greenhouse debate. The panel acknowledges the uncertainties in predicting the effects of man-made greenhouse gases such as carbon dioxide, methane and chlorofluorocarbons (CFCs), but nonetheless argues that it makes sense to begin certain inexpensive measures that will reduce the production of these gases — advice that runs counter to the current White House line of doing nothing that cannot be justified on grounds other than amelioration of the greenhouse effect.

On the other hand, the NAS panel stopped short of the environmentalist position that the possibility of global warming

demands a vigorous response now, and it takes two positions likely to irritate environmentalists. It calls for a series of measures to help mankind adapt to an increased greenhouse effect — such as research into farming under conditions of less water and more carbon dioxide — and even suggests that people in the United States could adapt to potential greenhouse changes with no more difficulty than experienced in adapting to the 'dust bowl' days of the 1930s. (One panel member, Jessica Matthews, vice-president of the environmentalist group World Resources Institute, wrote a sharp dissent from this view.)

And second, the NAS group advised research into 'geoengineering' techniques that could lessen the greenhouse effect, such as ways to increase cloud cover and thus reflect more sunlight back into space. Although the report emphasizes that it is calling only for research, not active measures, and that any such geoengineering would have to be done with extreme caution in light of potential unexpected consequences, the very idea of more meddling with the Earth's natural environment is anathema to many environmental organizations.

The report, which was released yesterday

(10 April), bases its conclusions on what it terms "conservative assumptions" on the possible global warming over the next 40 to 50 years — it posits a rise of 1–5 °C in average global temperature. The predictions of most global climate models fall somewhere in this range, assuming greenhouse gases increase to the point where there is effectively twice as much carbon dioxide in the atmosphere as in pre-industrial times.

Working from these assumptions, the Evans report states that "even given the considerable uncertainties in our knowledge of the relevant phenomena, greenhouse warming poses a potential threat sufficient to merit prompt responses." Fortunately, it continues, "the panel believes that substantial mitigation can be accomplished at modest cost. In other words, insurance is cheap."

The recommended insurance policy includes the following clauses:

- Reduce or offset emission of greenhouse gases by: continuing the phaseout of CFCs and developing substitutes that minimize or eliminate greenhouse gas emissions; studying the concept of 'full social-cost pricing' of energy, where the costs of various types of energy include such factors as their environmental effects, with the eventual goal of instituting such a system; increasing energy conservation and efficiency by such means as nationwide building codes and improved automobile gas mileage; planning future energy supplies with an eye towards greenhouse warming, so that natural gas and a new generation of nuclear reactors would provide a larger fraction of the nation's energy needs; and reducing global deforestation.

- Enhance adaptation to greenhouse warming by: doing agricultural research aimed at farming in an enhanced greenhouse world; making the nation's water supply more robust; building structures such as bridges, dams and levees with a margin of error that allows for the possibility of higher water levels in the future; and acting to prevent losses of biodiversity.

- Improve knowledge for future decisions by: continuing to collect the climatological data that is necessary for building and testing climate models; strengthening our understanding of which mechanisms have important roles in the climatic response to greenhouse gases, so that global climate models can be improved; improving weather forecasts in order to make it easier to adapt to possibly more severe weather conditions; conducting research on how different species of plants and animals would react to global warming; and studying the social and economic aspects of global climate change.

- Evaluate geoengineering options to understand the advantages and disadvantages of various methods of artificially decreasing the greenhouse effect.

- Participate in international programmes to slow population growth and take part in other programmes aimed at an international response to global climate change.

Robert Pool

Contrary agenda for greenhouse research

Washington

EVERYONE agrees that more research is needed to understand just how the increasing amount of greenhouse gases in the atmosphere will affect the Earth's climate, but is some potentially valuable research being ignored for political reasons? A group of climate researchers suggests that it is, and has offered a set of research proposals aimed at correcting some of the supposed omissions.

"There is significant research both going on and being proposed by people who are not convinced that there is a looming cataclysm," says meteorologist Richard Lindzen. But, he adds, researchers who do not accept the hypothesis that global warming is threatening to cause catastrophic climate changes find it much more difficult to get their proposals funded. So Lindzen and 27 other scientists met last October at Arizona State University to suggest global change research not predicated on that assumption.

The product of that meeting, *Global Climate Change: A New Vision for the 1990s*, was released last month, and seems to have presaged many of the rec-

ommendations of the new NAS global climate change report (see above).

Several researchers, for instance, suggest studies to get more accurate data on past and present climate changes, such as a proposal by David Aubrey of Woods Hole Oceanographic Institution to measure sea-level rise with new, very accurate technology. Others want to take a close look at how increased concentrations of carbon dioxide will affect plants and ecosystems. Still others wish to look at ways to lessen greenhouse warming, such as a proposal by George Kukla at Lamont-Doherty Geological Observatory and three others to examine how industrially produced aerosols may help to cool the climate.

The group of proposals offers a research agenda that is "complementary" to the current line, says Patrick Michaels at the University of Virginia, a well known greenhouse sceptic. Although the package, which was sent to the White House, has yet to receive any official response, Michaels predicts that it will have a better chance of getting attention than individual complementary proposals.

R.P.

Japan chases space market

Tokyo

JAPAN is hoping to grab a slice of the multi-million dollar market for the commercial launch of large international telecommunications satellites, a market currently dominated by Europe's Ariespace company.

At the end of last month, the Science and Technology Agency (STA) established a government panel to examine transfer of state-owned space technology to the private sector. The panel was set up at the urging of the National Space Development Agency (NASDA), a semi-governmental organization affiliated to STA, and Rocket System Corporation (RSC), a consortium of 75 private companies established last year to handle procurement and manufacture of Japan's next-generation H-II rocket, which is being developed by NASDA (see *Nature* 345, 191; 1990).

RSC is hoping to make a bid for the launch of huge next-generation telecommunications satellites that will be operated by the International Telecommunications Satellite Organization (INTELSAT) from the mid-1990s onwards. Last October, INTELSAT approached NASDA about the possibility of using the H-II to launch two INTELSAT VII-A satellites in 1994 and 1995. The matter was referred to RSC because NASDA cannot legally engage in commercial activities. But RSC cannot make a bid for the launches until the Japanese government determines its position on technology transfer — hence the need for the STA committee.

The committee, headed by Shigeo Kobayashi of Tokyo Metropolitan Institute

of Technology, will look into the transfer of NASDA-developed rocket technology to RSC and will also examine issues associated with the use of NASDA's launch facilities for the H-II at NASDA's space centre on Tanegashima island in southern Japan.

Ultimately, RSC wants to take over management of the facilities on Tanegashima. But there are a number of legal hurdles. If, for example, an accident occurs during a launch carried out by NASDA, the Japanese government is legally obliged to pay compensation. But no such laws exist to cover launches by RSC.

For the time being, RSC and STA officials expect that RSC will entrust any commercial launches to NASDA in return for a non-profit fee. And one of the jobs of the STA committee will be to examine the costs involved in using NASDA launch facilities.

Hirofumi Osawa, chairman of the board of RSC and former president of NASDA, says that INTELSAT is 'very interested' in using the H-II for its next-generation satellites. Only Ariane rockets and the H-II will have the necessary power to place the heavy INTELSAT VII-A satellites in orbit, Osawa says.

Ariespace dominates INTELSAT launches with five out of eight orders for the next few years (the other three going to US Atlas rockets), according to figures provided by Ariespace's Tokyo office. But INTELSAT is keen to find alternatives and is actively encouraging RSC to submit a proposal, Osawa says. Furthermore, Japan is one of the biggest contributors to INTELSAT through the international telecommunications company Kokusai Denshin Denwa, and thus Japan has strong political leverage in the assignment of INTELSAT launches.

RSC's 31 staff members are busily preparing for a request for a proposal from INTELSAT that they expect to receive in August — by which time they hope the STA committee will have reached some decisions on the issue of technology transfer and the use of the Tanegashima launch facilities.

One of RSC's biggest headaches will be setting an appropriate price for H-II launches, says Hiroshi Imamura, vice-president of RSC. Launches of present-generation INTELSAT VI satellites typically cost about \$100 million each, according to Jean-Louis Claudon head of Ariespace's Tokyo office. When development of the H-II began in the mid-1980s, NASDA aimed for a target launch price of ¥8 million per kilogram of satellite — a very competitive price at that time — but the yen has since doubled in value against the dollar and H-II launch costs are now expected to be about 20 per cent higher than Ariane's, Osawa says.

RSC officials thus hope to make reliability rather than price their strongest selling point. Imamura proudly displays NASDA's 100 per cent success rate in the launch of 22 satel-

lites. In contrast, Naotaka Oki, director of STA's space development division, points to the 'incredible' accident last year when a carelessly left rag apparently caused destruction of an Ariane rocket carrying a payload of Japanese commercial satellites (see *Nature* 344, 695; 1990).

But RSC faces serious problems with the H-II rocket itself. NASDA is experiencing difficulties in developing the main engine for the liquid fuel rocket and its first launch has already been delayed a year until 1993. Any further delays would virtually rule out any chance of the H-II being used to launch an INTELSAT satellite in 1994 and RSC officials are not prepared to predict whether the H-II will be launched on schedule.

In the long run, another barrier stands in the way of RSC. Launches of NASDA rockets from Tanegashima and from the nearby space centre of the Institute of Space and Astronautical Science are limited to two 45-day windows each year in summer and winter by an agreement with a small band of local fishermen. This effectively limits NASDA and RSC to at most four H-II launches a year — hardly enough to run a commercial operation.

The fishermen claim the rockets might damage their tuna fisheries, and in return for permission to use the 90-day window, the institute pays them ¥400 million (about \$3 million) a year for 'positive' improvement of their fisheries operations, Oki says. (It should be borne in mind that a very powerful member of the ruling Liberal Democratic Party held the Tanegashima constituency at the time the agreement was made.) Before RSC can become a strong competitor in the commercial space business, a new agreement will have to be reached with the tuna fishermen of Tanegashima. **David Swinbanks**

SPACE SCIENCE

ESA and Hungary cooperate

Munich

THE European Space Agency (ESA) this week signed a cooperation agreement with Hungary, its first such with an Eastern European country. The agreement provides for cooperation between ESA and Hungary in space science, Earth observation and environmental protection, microgravity and telecommunications.

The agreement was made on the basis of 'non-exchange of funds', according to Simonetta Cheli of ESA in Paris, but ESA will nevertheless provide free consulting in areas such as telecommunications and improve access to databases.

The Soviet Union signed a similar agreement last year, and Poland, Czechoslovakia and Romania have all expressed interest in cooperating with ESA. But full or associate membership in the 13-member organization, which would involve payment, is a long way off. Steven Dickman

Blurred lines

Tokyo

HIROYUKI Osawa, chairman of the board of RSC, provides a classic example of how in Japan the division between government and industry can become quite blurred.

RSC officials insist that RSC is a private organization that cannot directly influence government decisions. Similarly, Naotaka Oki, director of STA's space development division, is at pains to emphasize that the new STA committee established to look at space technology transfer is not directly concerned with the commercialization of Japan's space industry or RSC.

But their protestations seem rather unconvincing when one looks at interrelations between the different entities. Osawa is a former vice-minister of STA and former head of NASDA. Furthermore, while serving as head of RSC, he is also a member of the government-run Council of Science and Technology with an office four floors below Oki in STA. It seems certain that Osawa will be able to help smooth out any legal and government barriers that lie in the way of commercialization of Japan's space industry. **D.S.**

Successful satellite lift-off

Washington

AT 15 tonnes, the Gamma Ray Observatory (GRO) is the heaviest scientific satellite ever carried by the Space Shuttle. It was launched last Friday (5 April) and deployed two days later from the cargo bay of space shuttle Atlantis into Earth orbit, where it joins the Hubble Space Telescope as the second of four 'Great Observatories' planned by the National Aeronautics and Space Administration (NASA). The third and fourth Great Observatories are the Advanced X-ray Astronomy Facility (AXAF), tentatively scheduled for launch in 1998, and the Space Infrared Telescope Facility (SIRTF), not yet approved but endorsed by the recent Bahcall committee report as the highest priority among big astronomy projects (see *Nature* 350, 185; 21 March 1991).

GRO comprises four separate instruments, each designed to study the gamma-ray sky with different spatial resolutions and in different energy ranges. The unusual weight of GRO is due to two of its instruments in particular, EGRET (Energetic Gamma Ray Experiment Telescope) and COMPTEL (Imaging Compton Telescope),

GERMAN RESEARCH

Berlin finally licenses reactor

Munich

WITH the recent licensing of the refurbished reactor at the Hahn-Meitner Institute (HMI) in Berlin, German and foreign researchers can at last begin to use a nuclear reactor that serves primarily as a neutron source. The licence was announced by the newly elected Berlin government on 26 March. The reactor will begin a testing phase this week, and should be available for experiments by the end of the year.

The decisive factor in licensing the reactor was the resurgence in Berlin of the conservative Christian Democratic Union or CDU. CDU, which made licensing the reactor a high priority, received the most votes in elections on 2 December 1990.

The licensing had been delayed for nearly two years because of opposition from the local Green Party, known as the *Alternative Liste*. In 1989, the Greens halted the licensing procedure, begun in 1979, as it neared a conclusion. They objected to licensing the 10-MW reactor on the grounds that no suitable arrangements had been made to dispose of the waste it generated (see *Nature* 346, 688; 1990).

Steven Dickman

both of them the product of US-European collaborations. Because gamma rays cannot be focused by refracting devices such as optical lenses, imaging in these telescopes is accomplished by means of 'grazing incidence' reflectors, in which gamma rays are made to skate almost tangentially across a smooth metal surface, being deflected just

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Preparing for launch — the observatory undergoes tests.

enough to bring them to a focus a metre or so away. To produce a useful photon collection area, a series of slightly conical cylinders is assembled concentrically to form a massive gamma-ray 'lens'.

For the first 15 months of GRO's mission, both EGRET and COMPTEL will be devoted largely to sky-survey work. More sensitive by a factor of ten than any previous gamma-ray detectors, they are likely to trawl up large numbers of new astrophysical gamma sources, as well as measuring accurately the cosmologically important gamma-ray background.

Completing GRO's line-up are the Oriented Scintillation Spectrometer Experiment (OSSE) and the Burst and Transient Source Experiment (BATSE). OSSE will measure gamma-ray spectra in the energy range 100 keV to 10 MeV, where characteristic line emission from nuclear processes and radioactivity is produced. Its main observational targets will be astrophysical objects such as supernovae and novae, where heavy elements are manufactured, and pulsars and X-ray binaries, which generate electrons and photons energetic enough to induce nuclear activity.

BATSE consists of eight identical instruments mounted at GRO's corners and designed to watch the sky constantly for gamma-ray bursts. First observed two decades ago, gamma-ray bursts have been catalogued by the hundred but remain largely unexplained; they occur sporadically and unpredictably, and have not been conclusively associated with any known astrophysical object. BATSE's eight detectors can see almost the whole sky, and when a burst is seen by one of the units, as many as three others can quickly be trained in the same direction.

David Lindley

Ozone loss worse than expected

Washington

THE protective ozone in the Earth's atmosphere is disappearing much faster than previously thought, according to a study done by researchers at the National Aeronautics and Space Administration (NASA). The results, which have been reviewed at a workshop but not yet submitted for publication, were announced last week by William Reilly, director of the Environmental Protection Agency.

"These data suggest depletion of four to five per cent has occurred since 1978 over the United States," Reilly said. "Past studies had shown about half that amount." The ozone depletion is widely believed to be caused by the release into the atmosphere of chlorofluorocarbons (CFCs), man-made gases used for refrigeration, the manufacture of insulating foams and as cleaning solvents in making computer chips.

Besides the larger-than-expected drop, ozone loss has been found further south than previously seen. And more importantly, the ozone depletion has been detected well into the spring months, whereas researchers had previously believed that ozone depletion was significant only in winter. (Warmer temperatures are believed to slow down the chemical reactions by which chlorine from the CFCs breaks down the ozone.) Because people are more likely to be outside and exposed to the Sun in the warmer months, and because scientists estimate that for every one per cent decrease in upper atmosphere ozone an extra two per cent of the Sun's ultraviolet radiation reaches the Earth's surface, the springtime loss threatens to increase the incidence of skin cancer. Reilly said that the ozone depletion could nearly double the skin cancer rate over the next 50 years, causing as many as 200,000 extra deaths in the United States alone.

The data on which these conclusions are based were taken by the 'total ozone mapping spectrometer', an instrument carried on a NASA satellite. The instrument measures 'column ozone' — the amount of ozone in the atmosphere above a given spot on the Earth's surface. The data were analysed by Richard Stolarsky, an atmospheric scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

The ozone depletion extends across much of the Northern Hemisphere above a latitude of 30–35 degrees, including most of the United States, Europe and Asia. It may bolster demands to phase out the production and use of CFCs even more quickly than agreed in the 1987 Montreal Protocol, which calls for industrial nations to cease production by 2000 and less developed nations by 2010.

Robert Pool

Canberra squeezes ANU

London

THREE months of threat and counter-threat over the future of the Australian National University (ANU) at Canberra have been ended by a compromise announced two weeks ago by the federal minister of higher education, Peter Baldwin.

But researchers at ANU are up in arms at the proposals, and are hoping to appeal to the government over the head of the Department of Employment, Education and Training. Academics at other institutions, by contrast, tend to the belief that Baldwin's compromise is the best that ANU can hope for.

The argument revolves around the special role of ANU, created in 1946 both as a centre of excellence in research and as an indigenous means of training graduate students. ANU also incorporates a general university for undergraduate teaching (known to insiders as 'the faculties') which is small by Australian standards, with fewer than 6,000 students.

ANU researchers fear that research will be disproportionately damaged in the short run by a proposal that a percentage of research funds should be spent on externally approved collaborative research, and that the research schools' cohesiveness will be undermined by the five-year external reviews of the separate institutes' research that are now planned. And they are alarmed that the international standing of the institute will be harmed.

From the outset, research at ANU has been financed directly by the federal government, which spends more than A\$200 million a year at ANU, reckoning that A\$130 million of that is for research support. With the rapid growth of the Australian university system in the past 20 years, and the growing importance (in the past ten) of the Australian Research Council (ARC) as a source of research funds, ANU seems to have enjoyed a privileged access to funds.

Spurred by this lack of uniformity, the government last year appointed a committee under Sir Ninian Stephen, an ex-governor of Australia, to recommend an administrative framework for the Institute of Advanced Study, as the seven research schools are known collectively.

The committee's report was broadly supportive of the institute's present function, but recommended that a proportion of the institute's budget (5 per cent this year, rising to 12 per cent in 1998) should be earmarked for collaborative research with other Australian universities, that tenured appointments at the institute should decline to a third of the total by the end of the century and that there should be a further review of ANU in ten years.

Among other things, Stephen recommended that the John Curtin School of Medical Research, while remaining part of the academic structure of the Institute of

Advanced Study, should be provided with funds by the National Health and Medical Research Council. Baldwin accepts this proposal, which will take effect immediately.

By failing to recommend a separation between the institute and the undergraduate part of ANU, the Stephen report disappointed the federal government's declared fondness for university amalgamations (see *Nature* 343, 203; 1990). Prominent among these was a proposal that the faculties at ANU should be merged with the University of Canberra, until last year the Canberra College of Advanced Education.

The row over the future of ANU has been precipitated by the publication by the Department of Employment, Education and Training of an internal report by Ian Chubb, until last year pro-vice-chancellor at Wollongong University and now chairman of the Higher Education Council. Chubb's plan, which argued for splitting ANU, amalgamating all but the institute with the University of Canberra, has been hotly if predictably rejected by the council of ANU.

Baldwin's compromise leaves ANU intact, at least on paper, but the minister says that to remove "unproductive duplication and destructive competition" between ANU and the University of Canberra, each will in future have to develop a separate "educational profile" which must then be "jointly negotiated" with the federal government. But in his statement of 27 March, Baldwin said that amalgamation had "logic" in its favour, and urged the two universities to study the benefits of becoming a federal university.

On research, Baldwin proposes that 5 per cent (rising to 12 per cent) of the present budget should be transferred to collaborative projects, to be monitored for suitability by ARC. ANU academics argue that, with salaries accounting for 70 per cent of total costs, the effect on disposable research income will be much greater, reaching a half at the end of seven years.

Further ahead, each research school will be required to produce a research plan, and its progress will be assessed five years from now by a mechanism to be agreed with ARC; funds allotted to research schools then found not to be "performing" will be redirected through ARC. But ANU researchers say that merely by deciding to deal with each research school separately, the minister has robbed the Institute of Advanced Study of flexibility in redirecting its research.

Speaking from Sydney on Monday, Baldwin defended his proposals by saying that "surely it is reasonable" that ANU's research should be assessed externally, like other research. He also emphasized that "I've given them five years" (the interval before the first planned review) in which to put their house in order.

Academics elsewhere in Australia are less than enthusiastic for ANU's case. Some

argue that ANU research is no longer uniformly outstanding in Australia. Others say that the research schools have neglected 'the faculties' by failing to enliven the teaching of undergraduates.

Kurt Lambeck, director of ANU's distinguished Research School of Earth Sciences, says that while performance in this respect is patchy, his own colleagues in the past two years have been teaching courses in both the physics and applied mathematics departments of the undergraduate university, partly in the hope of spotting bright graduate students. His chief regret about Baldwin's decision is that it seems to foreclose the opportunity of creating an outstanding university in Australia.

ANU's next move is likely to be to accept the parts of Baldwin's policy that coincide with the Stephen report, but to reject those going beyond it. But because the policy can be carried through administratively, without special legislation, its protests may easily be overridden.

John Maddox

US APPOINTMENT

Eradicator of smallpox moves

Washington

DONALD A. Henderson has been officially confirmed as the associate director for life sciences at the US Office of Science and Technology Policy. Henderson, who has served since 1977 as dean of the Johns Hopkins School of Hygiene and Public Health in Baltimore, Maryland, is best known for his role in the eradication of smallpox. From 1966 to 1977, he served as

Peter C. Howard

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Henderson — filling a vacancy.

chief medical officer of the Smallpox Eradication Program of the World Health Organization, directing some 700 advisers and 200,000 workers in 69 countries in wiping out the disease.

The Office of Science and Technology Policy advises the President on science policy and also coordinates science programmes among various executive agencies. The position of associate director for life sciences has been vacant for more than a year, since James Wyngaarden resigned.

Robert Pool

Cutting student costs

London

THE Universities Funding Council (UFC) has unveiled a new funding system that aims to expand the number of British undergraduates while driving down the cost of teaching each student. The new system, which comes against the backdrop of the University of Edinburgh's facing financial disaster because it has too few students for the size of its faculty, will encourage universities to take on large numbers of 'fees-only' students without UFC funding by promising to reward those that do so with more UFC-funded student places in subsequent years.

This latest funding plan replaces last year's abortive 'competitive bidding' system, which sought to reduce the cost of teaching each student by making the universities bid against one another for funded undergraduate places. The idea was that universities

LONDON ZOO

Attempts to avoid closure

London

LONDON Zoo is scheduled to present a rescue plan today (11 April) to the Department of the Environment in a last-ditch attempt to avoid closure. The plan involves giving up some of the zoo's larger animals and transforming it into a centre for conservation education.

The cost of the change-over is estimated at about £12 million, most of which would come from private sponsorship. The government will be asked to contribute £3-4 million towards the implementation of the plan, but previous experience suggests it will be reluctant to agree. It offered a once-only grant of £10 million in 1988, but has since given no further aid to the zoo, which last year ran up a £4.9 million operating deficit. The government's handling of the zoo's financial predicament contrasts sharply with its behaviour towards other London institutions. Both the Natural History Museum and the Royal Botanic Gardens, Kew receive state subsidies.

Now the zoo's 1988 grant is half gone, and rather than see this fund shrink further, the Zoological Society of London, which administers the zoo and leases the 47-acre (19-hectare) site from the Department of the Environment, may choose to divert it elsewhere. This would mean closing the zoo and transferring a proportion of its 8,000-strong stock of animals to its nearly self-supporting, 600-acre freehold site at Whipsnade, about 35 miles north of London. The fate of the rest is undecided, and might depend on the extent to which any further government subsidy can be used to buy time for suitable arrangements to be made with other zoos. Henry Gee

that were eager to expand would offer to teach students more cheaply. The system failed, however, when universities tried to circumvent the system by submitting the vast majority of their bids for more students at the maximum prices suggested by the UFC — in effect operating as a cartel.

The new system will be more difficult to subvert. Any university that refuses to take on more students funded only through the minimal teaching fees charged to local education authorities (£2,650 per science student in 1991-92) will find their plans for growth stalled. Because almost all the universities wish to expand, the UFC believes that most will decide to take on more fees-only students, increasing the proportion of British undergraduates taught without UFC funding from the current 12 per cent. Those universities recruiting the most fees-only students will be favoured when the UFC comes to allocate the extra 25,300 funded places it plans to distribute between 1992-93 and 1994-95.

Although the Committee of Vice-Chancellors and Principals (CVCP) has yet to meet to discuss the new system, most vice-chancellors reject the UFC's central assumption that teaching quality can be maintained while reducing the average cost of teaching each student. Universities have increased their efficiency as far as possible, they argue, and the accusations of poor financial management made by the government during the 1980s do not now apply.

But the vice-chancellors' desire to present the universities as efficient, well-managed institutions may be undermined by the revelation that unless drastic action is taken, the University of Edinburgh will have a £5.4-million deficit in 1990-91 — much worse than the £4 million deficit it had predicted previously. And this is in addition to an accumulated deficit that totalled some £5 million at the end of 1989-90.

Sir David Smith, Edinburgh's principal, admits that the university has been "living beyond its means for a number of years", and blames defects in accounting for the previous failure to recognize the true extent of the problem. Edinburgh's difficulties stem from the fact that it has only 10.1 students for each member of staff — the lowest ratio in the country. But the university has now agreed an emergency package of measures to increase teaching efficiency, under close scrutiny from UFC officials.

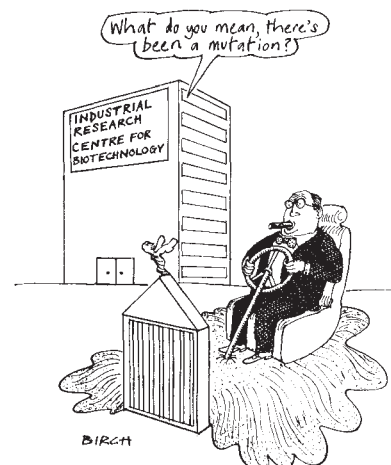
Smith says that Edinburgh will now take on ten per cent more students than it had planned in 1991-92, on a fees-only basis, and will continue a freeze on staff appointments introduced last year.

Edinburgh aims to break even within three years, but before then it may be forced to sell some £5 million worth of its £250 million assets, which are tied up in university-owned buildings and land. Edinburgh lecturers may

Car-eating bacteria

Munich

WHAT can be done with two million unwanted Trabants? Eastern Germans are rapidly scrapping the ubiquitous 2-stroke East German cars now that reunification has opened the borders to other automobiles, and they pose a problem in solid-waste disposal. But a group of German chemists believe they have a solution. Peter



Lietz and his coworkers at the Industrial Research Centre for Biotechnology in East Berlin are working on isolating strains of bacteria that will be able to digest the resins that make up the car's 650-kg body, reducing it to just 10 kg of residual biomass.

The East Berlin group holds patents on a strain of penicillium bacteria known as *Penicillium jantinnellum* that can break down the cellulose in the Trabant. They are now tackling the problem of breaking down the polyphenol resins that are another major component in the car. Once they have isolated an appropriate strain using conventional selection techniques or possibly genetic engineering, they will put it to work in a complex multistep process involving heat and mechanical breakdown as well as digestion.

Trabbits, as the cars are still affectionately known, will continue to be produced at the rate of 150 a day until the production plant in Zwickau is shut later this month.

Steven Dickman

also be asked this year to take a pay increase lower than that agreed nationally between the CVCP and the Association of University Teachers.

Edinburgh's admission of poor financial planning follows the finding by a University of Bristol financial committee that Bristol's unexpected £4.4 million deficit in 1989-90 was the result of poor financial control. The UFC has made Bristol's budget for 1991-92 provisional on the university's improving its financial management. **Peter Aldhous**

■ The CVCP this week publishes a booklet, *The State of the Universities*, which claims that the current pressure to expand while cutting costs will damage the universities irreparably.

Robocops: Stewart and Feder's mechanized misconduct search

Washington

"I DO, I admit, feel a little vindicated." After waging a five-year war against the science establishment, Walter Stewart is luxuriating in the spoils of victory. Even Ned Feder, his quiet, retiring *alter ego* and long-time partner, is feeling cocky. Their single-minded pursuit of Thereza Imanishi-Kari and Nobel laureate David Baltimore, two authors of what Stewart believed was a fraudulent *Cell* article, paid off last month when a National Institutes of Health (NIH) investigation reversed three earlier decisions and concluded that Imanishi-Kari had fabricated data.

It now appears that Stewart and Feder — who were attacked by senior scientists as they carried out their investigations — were right all along. For the researchers who criticized them it is a bitter pill to swallow. It may not, however, be the last. Science, it turns out, has not heard the end of Stewart and Feder.

Over the past six months, in the lull after the NIH took the reins of the Imanishi-Kari investigation, Stewart and Feder have been quietly planning their next assault on the research establishment. Tired of impossibly complicated cases that take years — and congressional intervention — to untangle, the two self-appointed defenders of scientific integrity went looking for something simple. They wanted to find cases of misconduct that

were so clear, so unarguably improper, that attention would be focused not on the details of the wrong-doing but on the culture that permitted it. They want to prove that science still has a misconduct problem — despite its protestations to the contrary.

The crime they settled on is plagiarism. And to find it, they have built what, for lack of a better word, might be called a misconduct machine. Using desktop computers and digital scanners, the machine reads scientific

The crime they settled on is plagiarism. And to find it, they have built what, for lack of a better word, might be called a misconduct machine.

manuscripts and compares them with an electronic library of research publications, looking for passages copied without attribution.

"This isn't misconduct of the patient-killing variety, but it's undoubtably misconduct," Stewart says. "We wanted to get a great test case, and plagiarism is something that everyone knows is wrong."

It now appears that they have already found a potential test case. Simply by feeding pages of books and review articles in related fields into their machine, they claim to have found at least one example of serious use of

borrowed and unattributed text. They have not, however, identified the scientist or scientists involved. To prove their point, they need to be able to name names and show how the system failed, but to do so may endanger their careers.

The reason is that the NIH may draw the line this time, refusing to allow them to use the case as a platform for another high-visibility assault on the science establishment. Although William Raub, deputy director of NIH, has decided to defend their misconduct investigations as legitimate study of the 'sociology of science' (see box opposite), he has insisted that they publish scholarly articles like other NIH scientists.

If they do find evidence of misconduct, they are expected to report the details to their supervisors (who would then presumably pass on the information to the NIH Office of Scientific Integrity for formal investigation). Like any other NIH scientists, they must submit their publications for internal reviews before submitting them to a journal. A congressional hearing, such as the four that Stewart and Feder orchestrated in their investigation of the Imanishi-Kari case, would not be considered an adequate publication, Raub says.

It remains to be seen how Stewart and Feder will reconcile this challenge. (After several initial interviews, they withdrew their cooperation on this article.) They would clearly like to document at least one ugly case of plagiarism, to prove that it happens and is tolerated by the scientific community. But they now know that publicly naming names or further freelance investigating may cause

If Orwell had had a Macintosh (and a scanner)

Bethesda

HACKED together with off-the-shelf Macintoshes and a tangle of cables and scanners, Walter Stewart and Ned Feder's new misconduct machine is an electronic bloodhound on the trail of plagiarism.

The heart of the beast is two Macintosh computers set up on a long counter, each with a matching scanner and a connection to a standard AppleTalk network. Stewart and Feder feed pages from books and articles into the scanners, where they are digitized to produce electronic images.

Commercial optical-character recognition (OCR) software scans the pages, translating the printed words into a computer-readable text file. Once the OCR software has been trained to read the particular typeface used in a certain book or article, it can understand more than 99 per cent of the text. Most remaining errors are caught by running the file through a spell-checker to spot misread or dropped characters.

After the pages are scanned, read and checked, they are saved on disk (the laboratory has several hard disks on

the network, amounting to a total of some 1.2 gigabytes of data space) to join the growing database of research articles and books that have become the test population for the plagiarism machine.

The heart of the actual search for copied text is what the two researchers call their 'comparison engine'. It examines two documents simultaneously, looking for passages that are identical or nearly so.

Special software, written by Stewart in the object-oriented C++ computer language, converts the text into a series of unique 'hash' numbers, each representing a 30-character passage. After assigning the string a number, it moves the starting point forward five characters and reads the next 30-character string. Eventually, the full text is reduced to thousands of overlapping text segments, each stored as a single 32-bit number.

Another piece of Stewart-written software, called the 'bold-face tool', then compares the two sets of numbers, looking for matches. It creates new files consisting of the original text, with any identi-

cal passages highlighted in bold.

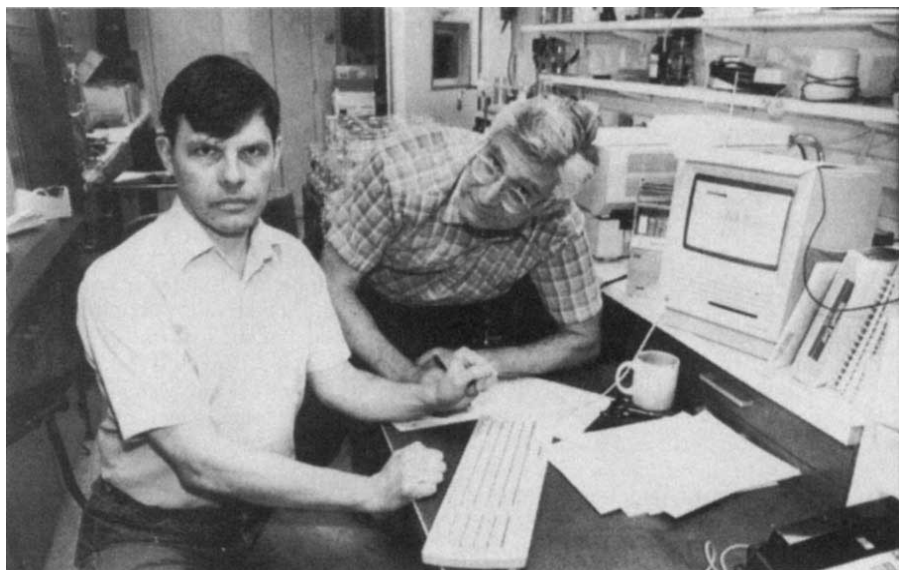
The researchers examine the highlighted passages in the context of the full text, to determine which are legitimate duplications — attributed to the original source, generic phrasing, or some other reasonable explanation — and which are potentially plagiarism.

"The typical finding is that there is a reasonable explanation," says Stewart. Yet they say they have already found some substantial passages that have been apparently copied without attribution or permission. Whether that represents real plagiarism or a harmless oversight remains to be seen.

Although the process requires a good deal of human supervision, the researchers have refined it to the point that they can digest about 500 pages a day — a typical book.

As far as Stewart and Feder are concerned, the machine is ready to go. All that remains in their misconduct experiment is for them to grab a stack of journals and books in some chosen discipline, fire up the computers and wait to see who has been naughty and who has been nice.

C.A.



AFP/Marty Katz/jamg

Stewart and Feder — in pursuit of the perfect misconduct case.

them to be dismissed.

This would not be the first time they have risked their careers to expose what they see as the flaws in science. To the NIH's vocal displeasure, Stewart and Feder emerged as fraud-busters in 1987 when they published an article in *Nature* on the "Darsee affair" (325, 207; 15 January 1987). John Darsee, a young research physician at Harvard with an unusually long string of papers to his name, had been found guilty by NIH of fabricating data in a cardiology experiment on dogs.

After that finding, Stewart and Feder took it upon themselves to review each of Darsee's works.

They soon discovered that some co-authors were nothing but "honorary", having had no direct part in the work, and that others did not even know their names were on Darsee's papers.

The Darsee paper was published only after some two years of negotiating between the pair, *Nature* editors and the inevitable lawyers. During that period, as they allowed their own biochemical research to lapse and began turning to the Imanishi-Kari case, the NIH started to put pressure on them to increase their scientific productivity (they have not published a basic research paper since 1983). Although NIH eventually rewrote their job descriptions to allow them to spend 20 per cent of their time investigating misconduct, for all practical purposes they had already become full-time fraudbusters — to the dismay of many of their colleagues at the NIH.

Stewart and Feder deflected some of the NIH pressure in 1989 by wangling an assignment to the congressional investigative committee of Representative John Dingell (Democrat, Michigan), who had taken an interest in the Imanishi-Kari case. But after a year of investigations and a hail of criticism from the scientific community, Dingell returned them last year, apparently deciding that they had become more of a liability than a help.

Back again at the NIH, the pursuit of the perfect misconduct case has taken over Stewart and Feder's life. The sub-basement NIH laboratory where they once studied neurology and genetics has become a virtual clearing-house of research wrongdoing. They officially stopped doing science eight months ago, although they admit that misconduct has been their only real research subject since 1988. The snails that they had so carefully bred since first coming to NIH were finally tossed out last year to make room for walls of new filing cabinets full of evidence. The hood where they years ago mixed chemicals for dye experiments now holds vertical folders.

Elsewhere at NIH, rumour of Stewart and Feder's new project is raising troubling questions. Should NIH be using taxpayers' money to support the investigations of two self-made detectives when it already has a misconduct office? Can Stewart and Feder be trusted to play fair, and to search for all possible explanations for what might appear to be plagiarism before blowing the whistle? As one NIH researcher puts it, the idea of Stewart and Feder feeding random literature through their machine in a search for wrongdoing "is practically Orwellian".

Stewart argues that the pair are saving science from itself. If, by documenting relatively minor offences, they can point out the flaws in the established self-policing method of ensuring scientific integrity, they can help to avoid worse scandals in the future. "It all depends on the first case we get out of the box," he says. "If it's big enough, it will make a difference." But what might happen if that case should not turn out to be a clear case of misconduct — or worse still should turn out to be an outright mistake — goes unsaid.

Despite the victory in the Imanishi-Kari case, an untested misconduct machine is dangerous at any speed. With the power to ruin careers, even a test-drive could cause disaster.

Christopher Anderson

NIH: working the ethical edge

Bethesda

WALTER Stewart and Ned Feder may work for the NIH National Institute of Diabetes, Digestive and Kidney diseases (NIDDK), but their similarities with other NIH scientists end there. Although the two researchers have a laboratory and nominal scientific responsibilities, they have essentially ignored their research assignment for the past three years, as they turned their attention exclusively towards misconduct.

Their situation as self-appointed internal investigators has understandably raised eyebrows, both on the NIH campus and elsewhere. Some critics have accused the NIH of tolerating Stewart and Feder's freelancing only because they feared the congressional investigative committee of Representative John Dingell, for whom Stewart and Feder worked for much of last year.

In an interview, William Raub, who was until this week the acting director of the NIH, explained why the agency continues to allow Stewart and Feder to chart their own course:

"We're working the edge of what we think is the intellectual freedom envelope. We recognize that it involves a lot of controversy, but we want our scientists to have the flexibility to do what they think they should for the progress of science. This place lives on ideas. We're loath to squelch ideas, even if they come with nettles.

"Until [last year], they said that the misconduct work was a purely transitory state. But it has become increasingly clear that are still intently pursuing misconduct. They are, however, careful to develop it as a work of scholarship.

"One might argue that who but scientists ought to be in the business of examining the fidelity of science. If their work — in the short term — sheds some light on the profession, then it may be useful. We have encouraged them to give it some focus — to concentrate on the sociology of science, not *ad-hominem* attacks on individual scientists. However, if they continue to do this for the long term, the director of NIDDK will find it hard to understand how it is in the interest of the institute.

"We've taken what we think is the moral high ground. If they can achieve a standard of scholarship in their work — by giving it an even-handed treatment using established data and no editorializing or *ad-hominem* attacks — that would be uncomfortable but acceptable. We ought to be prepared to accept the truth.

"It would bother me if I thought they were playing 'amateur investigator' — I hope they're not. But there's no doubt about the intelligence and integrity of these guys and I have a certain affection for smart people."

C.A.

The arrow of time

SIR — Huw Price's efforts to "demolish one of the main positive suggestions" in our book *The Arrow of Time*¹ are flawed. In attempting to refute our arguments about the existence of entropy in chaotic systems, Price states that even if chaotic evolution destroys time-symmetric determinism, "what it leaves might well be time-symmetric indeterminism". But for highly unstable dynamical systems (including K flows and Bernoulli automorphisms, which are of considerable theoretical importance), the temporal symmetry is broken^{2,3}.

It is precisely because of this time-asymmetric indeterminism that one can define an entropy-like quantity (or Liapounov variable) which has the property of monotonic increase with time. Moreover, the conditions under which such an entropy exists can be formulated rigorously in the language of ergodic theory in terms of intrinsic properties of a dynamical system^{2,3}.

This approach avoids the problems associated with coarse-graining and initial condition arguments, which lead to 'entropies' that can decrease with time.

PETER COVENEY

Schlumberger Cambridge Research,
PO Box 153,
Cambridge CB3 0HG, USA

ROGER HIGHFIELD

The Daily Telegraph,
London E1, UK

1. Price, H. *Nature* **348**, 356 (1990).
2. Covey, P. V. *Nature* **333**, 409 (1988).
3. Goldstein, S. et al., *J. Stat. Phys.* **25**, 111 (1981).

PRICE REPLIES: My original objection rests on the following logical point.

Let T be any theory whose purpose is to describe the temporal evolution of physical systems of some kind. Thus when applied to a model comprising a physical system *s*, T takes as input the information that *s* is confined to a certain region of its phase space at a time *t*₀. It generates as output information about the state of *s* at any time *t*₁, where *t*₁ > *t*₀ in the assumed temporal coordinates of the theoretical model.

The crucial question is whether, in order to yield a corresponding prediction about a real system, T needs to take into account whether *t*₁ is actually later or earlier than *t*₀, in the ordinary time sense of the real world. If not, then T yields the same result in either case, and hence does not give rise to any temporal asymmetry. If, on the other hand, T does depend on this further information, then this can only be because at some point it already treats the past and future orientations differently — otherwise the extra information would be redundant. Hence in this case any resulting asymmetry will effectively have been built into the theory from the beginning. So in neither case do we have 'sym-

metry-breaking'; that is, a temporal asymmetry which arises where there was none before.

The argument is independent of the mathematical form of the theory concerned.

HUW PRICE

Traditional and Modern Philosophy,
University of Sydney,
Australia 2006

Humane killing

SIR — In recent years, a better understanding of the animal, particularly in the physiological and ethological fields, has placed a greater emphasis on accountability and ethics with regard to the use of animals in experiments. The development of various methods of anaesthesia and humane killing which do not affect the data for an experimental study involving animals is therefore of the utmost importance. A number of effective methods exist^{1,2}, but some of these affect the parameters of a study, are costly, not readily obtainable or simple distasteful.

I would like to draw attention to an alternative method of killing mice and rats that involves the intraperitoneal injection of 70 per cent ethanol^{3,4}. The effects of 70 per cent ethanol on the central nervous system of mice and rats are similar to that of carbon dioxide. Upon injection, ethanol affects the acid-base balance by lowering blood pH and plasma bicarbonate concentration and increasing plasma lactate and ketone bodies. These changes result in respiratory and metabolic acidosis and cause depression of the central nervous system. This method is humane, inexpensive and very simple. In instances where barbiturate drugs would normally be used, such as vaccine candidate evaluation, 70 per cent ethanol may be of immense value as an alternative, particularly in developing countries where barbiturate drugs have limited availability. Physical methods such as dislocation of the neck might be used but requires expertise and is often distasteful to the researcher.

In the light of many recent articles published on animal experimentation^{5,6}, a continuing concern for accountability with animal use will undoubtedly see the development of alternative and improve methods of anaesthesia and killing. Ethanol may be the first of many new techniques.

ROGER LORD

Malaria Vaccine Group,
The Queensland Institute of Medical
Research,
Bramston Terrace,
Herston, Queensland 4006,
Australia

1. Green, C. J. *Animal Anaesthesia* (Spottiswoode, Ballantyne, London, 1979).
2. Blackshaw J. K. & Allan, D. J. *Principles of Laboratory Animal Management*, Australian Society for the Study of Animal Behaviour (University of Queensland, Australia, 1985).
3. Lord, R. *Aust. Vet. J.* **66**, 268 (1989).
4. Lord, R., Jones G. L., Spencer, L. *Res. Vet. Sci.* (in the press).
5. Good, R. A. *Science* **248**, 538 (1990).
6. Hollands, C. *Nature* **339**, 248 (1989).

Gulf oil spill

SIR — In his report on the Gulf oil spill caused by the Iraqi forces¹, Peter Aldhous discusses the suitability of recent techniques of bioremediation for tackling this catastrophe. Bioremediation technology involves spraying beaches and floating oil with cocktails of oil-feeding bacteria and/or nutrients that stimulate indigenous oil-feeders. The Texas company Alpha Environmental and its UK sister Alpha Biological Treatment Services are said to be standing by to move tonnes of bacterial cocktail to the Gulf.

Our group has been involved in research on oil-degrading microorganisms in Kuwaiti soils and coastal waters. We have found that one bacterial genus, *Rhodococcus*², predominates in the Kuwaiti (probably in the entire Gulf) environment, counting up to tens of millions of cells per gram of soil. Kuwaiti strains, which are adapted to the extreme environmental conditions of the region, proved much more active in oil degradation than strains isolated elsewhere. From the ecological point of view, the indigenous strains would severely compete with strains of the bacterial cocktail, and the oil-feeders could thus become antagonistic.

During oil degradation, bacteria frequently excrete specific detergents that may dramatically inhibit other oil-feeding strains³. One may thus conclude that the most suitable approach would be the spraying of nutrients to stimulate indigenous oil-feeding strains. Enhanced oil degradation would probably then follow, but there is reason to believe that new environmental hazards would also arise. *Rhodococcus*, like many other oil-feeding actinomycetes, excretes large amounts of trehalose glycolipids as detergents. These compounds show a degree of toxicity⁴, and in addition, frequently contain mycolic acids, whose relation to pathogenesis is known⁵. In this context, it should be remembered that the Gulf is the main source of drinking water in this region.

With regard to the volume of oil released (about 1.5 million barrels), which turned out to be considerably less than at first assumed (about 10 million barrels), the concerned authorities might be better advised to tackle this catastrophe using the more tedious but safer physical means of oil removal from water.

SAMIR S. RADWAN*

Department of Botany and
Microbiology,
University of Kuwait,
PO Box 5969, Safat 13060, Kuwait

1. Aldhous, P. *Nature* **349**, 447 (1991).
2. Sorkhoh, N., Ghannoum, M. A., Ibrahim, A. S., Stretton, R. J. & Radwan, S. S. *J. appl. Bact.* **69**, 856-863 (1990).
3. Foght, I. M., Gutnick, D. L. & Westlake, D. W. S. *Appl. environ. Microbiol.* **55**, 36-43 (1989).
4. Jardine, I., Scanlan, G., Mc Neil, M. & Brennan, P. *J. anal. Chem.* **61**, 416-422 (1989).
5. Toubiana, R., Berlass, J. Sato, H. & Strain, M. *J. Bact.* **139**, 205-211 (1979).

* Present address: Von-Einem-Strasse 25, D-4400 Münster, Germany.

Sad tale of an endangered zoo

Peaceful English contentment has been disturbed by rumours that the London Zoo may have to be closed later in the year, but neither the reasons for the sudden fuss nor the need for such a catastrophe are clear.

WHEN the world is full of people worrying about the future of other than the human species, how can the London Zoo be in danger of commercial failure? This week, the zoo has been putting it about that it may have to close later in the year, depending on the outcome later in the week of a further appeal for funds to the British government. At one stage, there was even a suggestion (quickly denied) that many of the 8,000 animals at the London Zoo might have to be killed, or 'put down' as the British say.

The most charitable explanation for this state of affairs is that animal-lovers are no longer willing to pay to see the objects of their affection locked up in cages, but that is at best a poor explanation, supposing as it does that the fault lies with the customer. It is easier to believe that the London Zoo, which has been struggling with deficits for years, has with equal consistency failed to attract its natural customers and to provide them with the incentive to return.

A zoo, for better or worse, is a commercial operation, which is not to say that it is simply a business. Making it possible for people who mostly live in Britain to see exotic animals is a public service, a way of reminding mostly urban dwellers of the diversity of the life beyond the limits of their subway networks and of the phenotypic inventiveness of the animal genome, appropriately shuffled. The difficulty is that, to keep animals in a zoo, it is necessary to feed them artificially and to pay keepers to do the work. That is where commercial reality intrudes on public service.

It is natural that most respectable zoos should have been founded by naturalists or zoologists, and that the search for commercial viability has often been painful and uncertain. Often, the search turns out to be fruitless. The London Zoo, which claims to be the oldest of its kind, seemed to have been permanently saved two years ago, when the British Department of the Environment provided it with a capital endowment of £10 million. But in a recession year, that turns out to be insufficient, and the begging bowl is out again. The London zoos — there are two, one in Central London and one at a country site called Whipsnade Park, thirty miles north — have let the painful search for viability drag on for too long.

What does London lack that San Diego has? The San Diego Zoo (see *Nature* 349, 457; 7 February 1991) is zoologically respectable, even distinguished in projects such as the return of the condor to the wild, has a financial turnover roughly ten times that of the London Zoo (excluding Whips-

nade), but derives only two per cent of its income from public sources (a property tax levied by the City of San Diego). One of the most significant items in San Diego's profit and loss account is the income derived from family memberships in the zoo. The largest is the income from the sale of goods from the zoo's array of shops.

San Diego stays more or less in financial balance — it has no choice — but in the most recent financial year, the city-centre London Zoo had an income of £4.2 million and made a trading loss of £1.9 million. For the British enterprise as a whole, which includes both zoo sites and the research institute called the Institute of Zoology, the trading income amounted to £6.9 million, the cost of running the operation was £11.8 million and the loss amounted to no less than £4.9 million. (To be fair, the services such as these in London are contracted out to operating companies, and only the nett amounts of profit and loss are shown in the zoos' accounts.) Even so, for every £1 taken in admission fees, the enterprise had to draw down its investments by nearly £0.70.

That sounds dangerously like carelessness, especially because the news of the deficit followed almost on the heels of public excitement at a proposal that the city-centre zoo should take in an extra slice of the city park in which it is located so as to enlarge itself. And the truth is that it is carelessness of a kind. Part of the present trouble is that the recession has cut people's willingness to pay admission fees and then to spend money on other things when they are inside one or other of the London zoos, but it is careless to behave as if such evil luck can be countered by seeking sponsorship from commercial companies which are likely themselves to be more directly exposed to recessionary forces than even the zoos' admission turnstiles.

The only sure remedy in such circumstances is to spend less, but the zoos' staff at both sites increased between 1989 and 1990 by ten per cent (and its total salary bill by 25 per cent). No doubt the increases were desirable, but would it not have been better to postpone them than to risk closing one or other of the sites? That may seem a harsh and short-sighted view, but it is what passes for commercial reality.

So what should be done? To the question what does London lack that San Diego enjoys should be answered, first, climate (which tempts people to picnic at the zoo) and, second, foresight and imagination (which has made it worthwhile to foster the membership scheme over a period of half a

century). But the question can also be answered when asked the other way round; London's climate may leave something to be desired, but the city-centre zoo does enjoy 42 acres of central London, a mere bus ride from the places thronged by tourists most times of the year? And it is at the centre of a conurbation of 20 million people, ten times as great as that from which the catchment area from which San Diego's family memberships are drawn.

The first need is that the zoos (or the Zoological Society of London, which owns them) should resolve to keep both sites going, however painful the economies required for the purpose. If necessary, the Institute of Zoology, which exists mainly for research on members of the animal collection, but which also has an important conservation and field programme, should be spun off to some other institution or, at the worst, closed altogether. Without the zoos, it would not (and perhaps could not) exist. Why keep it going when its continuation threatens the survival of the zoos on which its own long-term existence hangs to the tune of more than £1 million a year? That sounds rough talk, but is there a choice?

Keeping the city-centre site in being is essential if the zoos are ever to make the claim on the attention of the general public from which their core support must eventually be drawn, but that does not mean that the same complement of animals must remain there indefinitely. Fewer animals in the city zoo might enable a smaller staff to show them off more interestingly, and more in keeping with more enlightened views of how animals in captivity should be kept. The notion of a commuter zoo in which some animals are visitors from the distant suburbs should not be scorned.

The zoos must also learn more about the markets, whose downturn has caused such tribulation, than they seem to know at present. At present they do little to tell their potential customers of what exhibits are particularly interesting, and when. The potential benefits of a general membership scheme (not just for zoologists and their partners) have not been appreciated, but would handsomely reward sustained encouragement (even if that meant breaking down some of the barriers between professionals and the unqualified masses). And if the worst comes to the worst, the market would always allow the zoos to put their own operation out to commercial tender. It would be surprising if there were no takers, even on the most onerous conditions.

John Maddox

Another hit for gene targeting

Christopher Wright and Brigid Hogan

Any lingering doubts that homeobox (Hox) genes are important regulators of vertebrate embryogenesis should be dispelled by the paper on page 473 of this issue¹ describing mutant mice in which both copies of the *Hox-1.5* gene have been disrupted. In *Drosophila*, it is well established that *Antennapedia*-type homeobox genes orchestrate early development by determining the positional value of specific segments. Similar homeobox sequences have been highly conserved throughout evolution, and a major challenge is to ascribe functions to these Hox genes in vertebrates. Chisaka and Capecchi now report the use of homologous recombination in embryonic stem cells to produce homozygous *Hox-1.5*^{-/-} mutant mice. The results convincingly demonstrate that Hox genes regulate normal vertebrate development, although the precise mechanism by which regulation is achieved is still unclear.

Most vertebrate Hox genes are found in four arrays called Hox-1, -2, -3 and -4, representing a quadruplication of the prototypic HOM-C (ANT-C/BX-C) cluster of *Drosophila*. Diverged versions of particular Hox genes are represented in several different clusters; for example, in mouse there are two genes closely related to *Hox-1.5*, called *Hox-2.7* and *Hox-4.1*. Unanswered questions about them include not only the function of individual genes but also the reason for their clustering and quadruplication. Localization of Hox-1 and -2 RNAs in the mouse embryo hindbrain has revealed a colinear relationship between the anterior border of expression of a Hox gene and its relative position within the cluster (reviewed in ref. 2). In some cases, anterior borders of Hox gene expression are coincident with hindbrain rhombomere boundaries. The anterior limit of *Hox-2.7*, for example, is the junction between rhombomeres 4 and 5, while the cut-off for *Hox-1.5* lies somewhere anterior to the otic vesicle³. As far as we know, patterns of Hox gene expression are similar in the neural crest arising from the hindbrain neural tube and in the mesoderm initially flanking it^{4,5}. These observations support the idea that the fate of cells in different anterior-posterior regions of the neural tube, neural crest and mesoderm is determined by their 'positional address' based on different combinations of Hox gene expression.

One way to test this hypothesis is to generate homozygous null mutant mice by gene targeting in embryonic stem cells⁶. Chisaka and Capecchi have disrupted the *Hox-1.5* gene by inserting the gene for neomycin resistance into the DNA-binding homeo-domain, so that *Hox-1.5* can no longer activate or repress downstream 'effector' genes. It is possible that such a truncated *Hox-1.5*

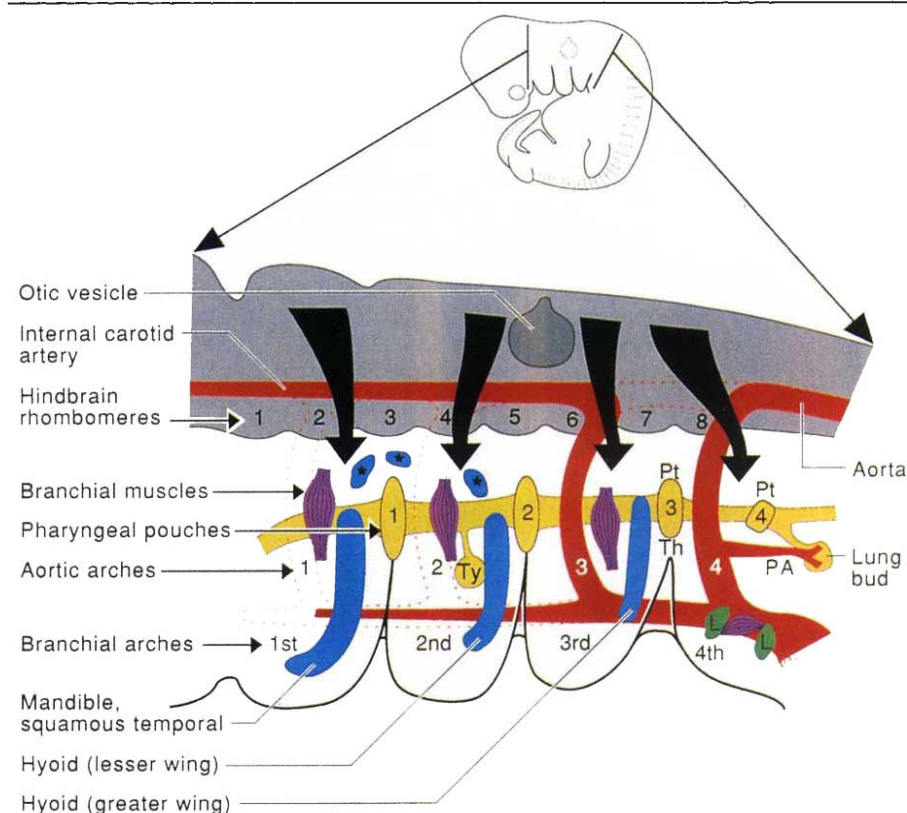
protein could interfere with endogenous protein-protein interactions, but the fact that the heterozygous mice are normal apparently rules out such a dominant-negative effect.

What phenotypes might result from genetic alterations in a vertebrate Hox gene? In *Drosophila*, ectopic overexpression of a HOM-C member usually causes the transformation of a segment into a more posterior one, while inactivation generally causes anterior transformations. For example, *Ubx* flies have an extra pair of wings. Abnormalities reported in transgenic mice ectopically expressing the *Hox-1.1* gene have been interpreted as posterior transformations of cervical somite identity⁷, implying that Hox genes function similarly in flies and mice. If this is so, one prediction is that inactivation of *Hox-1.5* would cause startling anterior transformations rivaling those seen in *Drosophila*.

This is not the case but, nevertheless, pro-

found alterations in development are found in the majority of the homozygous mutant mice, which survive for a few hours after birth. These defects include the absence of the thymus and parathyroids, a greatly reduced thyroid, and dysmorphologies of the heart and blood vessels. Moreover, the whole throat region is so abnormal that air enters the stomach, and the pups probably die from combined respiratory and circulatory failure.

As shown in the figure, the lower face and throat arise from paired segmental structures called branchial arches and pharyngeal pouches. The mesenchyme of the arches comes from two sources: mesoderm flanking the neural tube, and neural crest cells which migrate in large numbers from the dorsal surface of the hindbrain. This neural crest forms most of the connective tissue in the head and neck, while paraxial mesoderm produces only the voluntary muscles. Although the segmented organization of this region is obvious early in development, large-scale cell movements and mixing occur later, so that the rudiments of structures such as the thymus and thyroid migrate to positions far posterior of their



Segmented organization of the branchial arches of the mammalian embryo. The region shown in a nine-day mouse embryo is enlarged and highly schematized. Most of the mesenchyme of the branchial arches, including the mesoderm of the thymus, parathyroid and thyroid glands, the outer walls of the blood vessels, and the cartilages marked in blue, is derived from neural crest cells that emigrate from the hindbrain neural folds (arrows). Paraxial mesoderm gives rise only to voluntary muscles (purple). The cartilages of the larynx (green) are derived from lateral mesoderm. In the chick many endothelial cells of the blood vessels are derived from highly migratory angioblasts arising from paraxial and lateral mesoderm. The epithelium of the thyroid, parathyroid and thymus develops from endoderm (yellow). Middle ear ossicles (*); laryngeal cartilages (L); pulmonary arteries (PA); parathyroids (Pt); thymus (Th); thyroid (Ty). (Diagram produced by D. Noden.)

QUANTUM MECHANICS

Light and the single photon

Stephen M. Barnett

origin. In *Hox-1.5* null mutants, the defects appear to arise in structures derived from all the branchial arches and pharyngeal pouches. But because of the complex interactions between neural crest, mesoderm, ectoderm and endoderm in the development of the head it is impossible at present to attribute the primary defect to any one cell type. Indeed, this problem may be very difficult to unravel without recourse to the kind of grafting experiments done in chick embryos. A general pattern may emerge, however, as more *Hox* genes are mutated, and, as Chisaka and Capecchi suggest, this may lead to an understanding of human congenital defects in which branchial arches are affected, such as DiGeorge's syndrome.

Almost as interesting as the defects in the homozygous mice are the regions which are apparently normal, although they express *Hox-1.5* in the wild-type embryo. These include the hindbrain rhombomeres, cranial nerves, spinal cord, lung, stomach, spleen and kidneys. This suggests that in some regions of the embryo there is functional redundancy between *Hox* clusters. In other parts, however, particularly where the overlapping patterns of *Hox* expression are most clearly seen, individual *Hox* genes probably have unique functions in determining the developmental fate of specific cell types. One unanswered but important question is whether *Hox-1.5*, *Hox-2.7* and *Hox-4.1* are co-expressed in all cells in this region, or whether the patterns are subtly different. Also unclear is whether some of the mutant defects arise because the knockout of *Hox-1.5* alters the normal expression patterns of other *Hox* genes.

Partial functional redundancy may explain the results of other recent gene-targeting experiments. For example, homozygous *en-2* and *wnt-1* null mutant mice exhibit defects in only a subset of the tissues expressing the respective genes, both of which belong to gene families⁸⁻¹⁰. Given that other members of these families are now being targeted, including other *Hox* genes, we can expect further news about gene interactions during mammalian development in the coming months. □

Christopher Wright and Brigid Hogan are in the Department of Cell Biology, Vanderbilt University Medical School, Nashville, Tennessee 37232, USA.

QUANTUM theory was first developed for light, and light continues to reveal some of its most surprising aspects. The surprises become strongest when dealing with single photons, the individual quanta of light whose behaviour can be both particle-like and wave-like. Two new papers (one an experimental test and the other a theoretical prediction) have again focused attention on the bizarre properties of these fundamental excitations of the electromagnetic field.

In 1984, M. V. Berry revealed hitherto un-

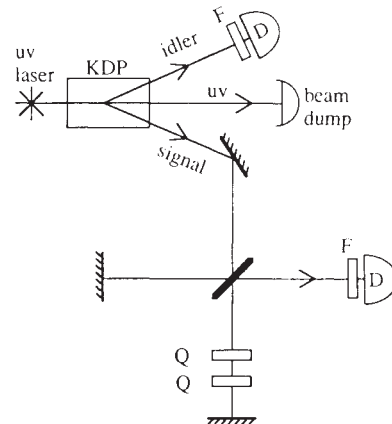


FIG. 1 Simplified scheme of Kwiat and Chiao's apparatus. The KDP crystal creates pairs of idler and signal photons. The latter enter a Michelson interferometer containing a half-silvered mirror (solid diagonal) offering a choice of paths. One path includes two quarter-wave plates (Q), one rotatable. Shaded rules, mirrors; F, filter; D, detector.

expected properties of the wavefunction's phase, a periodic, numerical factor responsible for interference effects in quantum mechanics (see M. V. Berry *Sci. Am.* **259** (6), 26–35; 1989). His remarkable discovery was that the smooth variation of the experimental parameters in an apparatus could lead to a wavefunction acquiring a phase factor additional to the phase associated with the normal dynamical evolution of the state. This additional phase would be accessible to experimental investigation whenever the parameters were returned to their original form, and so be a record of the changes the system had experienced.

Manifestations of Berry's phase were soon found in almost every branch of modern quantum theory. In addition to its quantum manifestations, two phenomena describable in the classical, electromagnetic picture of light were shown to have the same origins as Berry's phase. The first is the way the polarization of a light beam is rotated as it propagates through a helically wound optical fibre. The second, Pancharatnam's phase, is the phase acquired by a light beam on undergoing a sequence of polarization changes. P. G. Kwiat and R. Y. Chiao have now extended these classical manifestations of Berry's

phase firmly into the quantum regime by performing a demonstration of Pancharatnam's phase with single photons (*Phys. Rev. Lett.* **66**, 588–591; 1991).

Their experiment consists of a crystal which splits ultraviolet laser light into pairs of "signal" and "idler" photons. The signal photon passes into a Michelson interferometer in one arm of which, quarter-wave plates produce a sequence of polarization changes (Fig. 1). The probability that the photon will leave the interferometer and be detected will depend on the Pancharatnam–Berry phase induced by the polarization changes. The rate of coincidences between photodetection of the idler beam and the output from the Michelson interferometer in the signal beam oscillates as the angles of the polarizers in the interferometer are changed.

This observation is more than the reproduction of a classical phenomenon at the quantum limit of a single photon. It also contains an element of the intrinsically quantum phenomenon of 'nonlocality'. The visibility of the oscillations is determined by the relative sizes of the coherence length of the signal photon and the difference in path length between the two arms of the interferometer: but this coherence length may be modified by the bandwidth of (the range wavelengths passed by) a filter placed in front of the detector in the idler beam. Thus the visibility of the oscillations generated by single-photon interference in the Michelson interferometer is determined by the filtering of the spatially separated idler photon prior to its detection. This experiment is the first experimental demonstration of a nonclassical optical interference effect that includes a phase shift due to Berry's phase.

In their famous paper of 1935, Einstein, Podolsky and Rosen proposed that quantum mechanics may lack some necessary elements of reality and therefore constitutes an incomplete physical theory. It was not until the work of Bell that we had a means of deciding experimentally whether quantum mechanics is incomplete or whether it contains features that are incompatible with local realism: specifically, features that allow "spooky actions at a distance". So far, the experimental evidence has strongly supported quantum mechanics and the necessary conclusion seems to be that we must forego the comfort of local realism in favour of the non-locality of quantum mechanics.

The trial of quantum mechanics against local realism has concentrated on experiments, both of the type involving apparatus and of the *gedanken* or thought type, which involve pairs of correlated particles, described by a single 'entangled' wavefunction. But now Tan *et al.* demonstrate the possibility that only one particle may be required

- Chisaka, O. & Capecchi, M. *Nature* **350**, 473–479 (1991).
- De Robertis, E. M., Oliver, G. & Wright, C. V. E. *Sci. Am.* July, 46–52 (1990).
- Gaunt, S. J., Sharpe, P. T. & Douboule, D. *Development Suppl.* **104**, 169–179 (1988).
- Frohman, M. A., Boyle, M. & Martin, G. R. *Development* **110**, 589–607 (1990).
- Hunt, P., Wilkinson, D. & Krumlauf, R. *Development* (in the press).
- Capecchi, M. R. *Science* **244**, 1288–1292 (1989).
- Kessel, M., Balling, R. & Gruss, P. *Cell* **61**, 301–308 (1990).
- McMahon, A. P. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
- Thomas, K. R. & Capecchi, M. *Nature* **346**, 847–850 (1990).
- Joyner, A. L., Herrup, K., Auerbach, B. A., Davis, C. A. & Rossant, J. *Science* **251**, 1239–1243 (1991).

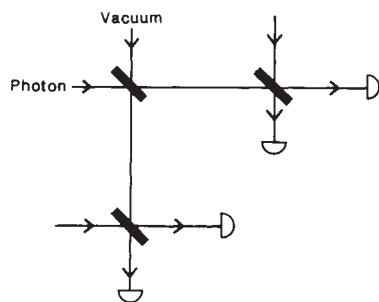


FIG. 2 The apparatus imagined by Tan *et al.* A single photon incident on a half-silvered mirror may be detected in either of two homodyne detectors (right and bottom), each including a local oscillator (unlabelled inputs). As described in the text, the 'vacuum' also incident on the half-silvered mirror must also be considered.

in order to demonstrate the nonlocality of quantum theory (S. M. Tan, D. F. Walls and M. J. Collett *Phys. Rev. Lett.* **66**, 252–255; 1991). It is still true that an entangled state light is required, but the insight of Tan *et al.* is that the entangled state may consist of single-photon states and the no-photon or 'vacuum' state. If a photon is incident on a semi-reflecting mirror, then the state of the outgoing field will be correctly described not by a photon that has been reflected or transmitted with probability 1/2 but by a superposition of the state corresponding to the photon being transmitted through the mirror and the vacuum reflected and the state corresponding to the photon being reflected and the vacuum transmitted. It may seem a little odd to think of the vacuum as an entity to be reflected or transmitted, but the vacuum is a state of the electromagnetic field and so must satisfy the correct boundary conditions at the surface of the mirror and may therefore be thought of as having a transmission and reflection coefficient.

EVOLUTIONARY BIOLOGY

A case of self-deception

R. I. Vane-Wright

"NATURE never deceives us; it is always we who deceive ourselves", said Rousseau in 1792. If he could have waited to hear Henry Walter Bates on the butterflies of the Amazon¹, perhaps he would not have been so sure. Bates introduced one of the most enduring notions of evolutionary biology — sheep in wolves' clothing, or "the deceptive way of life"². According to this idea, poorly protected organisms can gain increased immunity from predation by mimicking the appearance, smell, taste, feel or sound of well-protected species. The theory of mimicry has been potent: "the hypothesis that explains mimetic resemblance explains evolution" is a conviction for some³. But on page 497 of this issue⁴ we discover from D. B. Ritland and L. P. Brower that the viceroys butterfly (*Limnitis archippus*), for so long regarded as the classic example, can no longer be considered

The crucial physical point is that there is no element of physical reality associated with the trajectory of the photon. If we choose to count photons then we will find that either the photon has been reflected or it has been transmitted. However, if we measure the electric fields in the reflected and transmitted beams then we will find a correlation. This correlation is incompatible with the premise that the photon was either transmitted or reflected and therefore demonstrates the non-locality of the field at the single photon level. In the *gedanken* experiment of Tan *et al.*, the transmitted and reflected electric fields are measured by homodyne detectors (Fig. 2) and the phases of the local oscillators used in the detectors are local parameters. Changing one of these phases and making a measurement should change the measurement statistics of the other homodyne detector in a manner incompatible with local realism.

Single-photon sources of the type used by Kwiat and Chiao may make it possible to demonstrate single-photon nonlocality in the laboratory. The detection of the idler photon would alert the experimenter that a photon was on the way. This photon could be made to fall on a half-silvered mirror and homodyne detectors inserted in the transmitted and reflected beams. The problem would be that the experiment requires local oscillators of extremely low intensity. In fact a strong violation seems to require local oscillators in which, on average, less than one photon falls on the detectors during the coherence time of the nonlocal single photon. If this problem can be overcome, further insights may be gained into the surprising world of the single photon. □

Stephen M. Barnett is in the Department of Physics, King's College London, Strand, London WC2R 2LS, UK.

to be a batesian mimic of the chemically defended monarch (*Danaus plexippus*). So is batesian theory itself just a fabric of self-deception?

The theory of mimicry deals with the origin of a complex adaptation, involving genetics, ecology, physiology, behaviour and communication, with a clearly defined goal — convergence on a model signal pattern. As a result, opponents of darwinism have always tried to discredit Bates's idea. In defence, protagonists hedged the theory with many assumptions and qualifications. Although the idea survives (because it explains more than any alternative), it has been marginalized — mimicry is now often seen as little more than an intriguing evolution sideshow, part of "the genteel upper-middle-class fascination with snails and butterflies", as Richard Lewontin once put it.

One of the most telling points made against the theory is likelihood. If the selective value of chemical defence is so great that its influence can modify the appearance of other organisms, is it not more likely that an unprotected species will develop its own defences *de novo* rather than merely simulate the appearances of another? Bluff, with no stick to wield when challenged, is a risky sort of defence.

Undeterred, enthusiasts multiplied the number of cases to be explained by Bates's theory. This conviction was supported by an emergent myth: the idea that chemical defence in butterflies depends on sequestration of secondary compounds from larval hosts. It became axiomatic that only a few groups of plants (Asclepiadaceae, Solanaceae, Passifloraceae, Aristolochiaceae for example) could supply the necessary protection to confer model status. Similar-looking sympatric butterflies that did not feed on these plant groups were, *de facto*, batesian mimics.

This simple vision is now being strongly challenged^{5,6}. For example, the ithomiids, the main group of aposematic (warningly coloured) butterflies encountered by Bates, mostly feed on Solanaceae. But according to Brown⁷, "contrary to general wisdom . . . the Ithomiinae derive few or no defense compounds from their larval hostplants". Two other aposematic butterfly groups, the Acraeinae and Heliconiinae, largely feed on the Passifloraceae. But the cyanogenic glucosides which apparently protect these butterflies prove, at least in a number of cases, not to be sequestered from the cyanogenic compounds found in their passion vine hosts⁸. Indeed, the most potently cyanogenic *Acraea* species studied so far do not even feed on Passifloraceae⁶.

From observations such as these a new view is emerging — butterflies do not suddenly develop an ability to tolerate chemically protected hosts, shift to them, sequester their deterrents and then become aposematic, all in one seamless evolutionary sequence. Instead, as Ackery⁶ and others have proposed, an insect may first develop its own independent means of chemical defence. If this is sufficiently potent, as Miriam Rothschild has often pointed out, the insect is likely to become aposematic. New metabolic pathways of this type may also preadapt insects to shift to plants protected by similar chemical means. Having made this shift, it is possible that the insects then start to sequester or 'stock-pile' plant chemicals directly, as appears to be the case with at least some Acraeinae⁹, and with *Danaus* and asclepiad-derived cardenolides (although Brown *et al.*⁵ still favour the view that direct sequestration is likely to represent a primitive stage, citing the troidine swallow-tails and aristolochic acids as one of their examples).

Such host shifts followed by adaptive radiation could be responsible for the impression that the main aposematic groups are depend-

NUCLEAR PHYSICS

ent on plant secondary chemicals. However, if host shift and adaptive radiation do not always occur, we may also expect to find individual chemically protected species, closely related to less well protected sorts, and feeding on 'usual' foodplants. This appears to be the case with the viceroy, which feeds on the same hosts (notably *Salix* and *Populus*) as other North American butterflies of the same genus⁶.

The viceroy differs from its close relatives in two major ways: in the newly discovered chemical protection and in its dramatic colour pattern shift to look so similar to *Danaus plexippus*. Ritland and Brower⁴, using red-winged blackbirds in a novel bioassay, demonstrate that in Florida the viceroy, far from being poorly protected, is as unpalatable as the monarch, and significantly more so than another local *Danaus*, the queen butterfly (*D. gilippus*). Platt¹⁰, by hybridizing the orange viceroy with typical black-and-white *Limenitis* species, has created a remarkable series of intermediates between the presumed ancestral pattern and the monarch-like phenotype. Very strong selective value for the monarch pattern has been demonstrated in the field. The best description we can now give for the viceroy's life style is that of a distasteful, chemically divergent species strongly convergent on the warning pattern of the monarch.

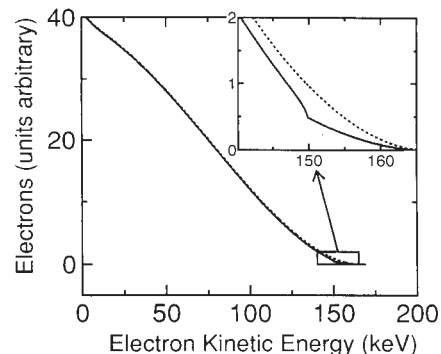
Batesian mimics, which may be much rarer than expected, are communication parasites. Müllerian mimics, which may include a number of individual species strongly divergent in signal pattern from their close relatives, are communication mutualists. Because, in addition to sharing the same warning signal, all members of a müllerian group are well-protected, it has been argued that no deception is involved and, therefore, they are not really mimics at all. But the wisdom of Rousseau's remark entreats us to keep anthropomorphism out of science. Nature does not deceive. The essence of mimicry lies in the biological concepts of parasitism and mutualism¹¹, not with human value judgements of 'deception'. The viceroy remains a classic example of mimicry — but not of batesian mimicry in particular. □

R. I. Vane-Wright is in the Biodiversity Programme, Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

Neutrino physics gets heavy

Harry N. Nelson

NUCLEAR physicists at Oxford have reported surprising evidence that one of the members of the family of neutrinos, the ghostly particles postulated by Pauli in his interpretation of nuclear β -decay, possesses a mass far greater than anticipated. The painstaking new experiment on the β -decay of sulphur-35 (^{35}S) by Hime and Jelley¹ apparently confirms previous hints² of this neutrino, which



The theoretical spectrum of electron energies in ^{35}S β -decay without 17 keV neutrino emission (broken line) and with 17 keV neutrino emission (solid line). To accentuate the effect, the probability for 17 keV neutrino emission is taken to be 50 per cent.

is so heavy that its existence would assault current views on cosmology as well as particle theory. Hime and Jelley's result finds support from two new experiments^{3,4}, performed at Lawrence Berkeley Laboratory in California, and the Ruđer Bošković Institute in Croatia, but the complete experimental case for this heavy neutrino is highly controversial, because earlier searches for it contradict the new experiments.

First hypothesized to explain the fate of energy apparently missing from nuclear β -decay reactions, neutrinos now constitute a fundamental class of elementary particles. Experiments at CERN recently proved⁵ that there are three types of light neutrino, associated in turn with the electron, the muon and the tau (the three charged 'leptons'). These neutrinos are assumed to have no mass (as well as no electrical charge) in the standard electroweak theory, which makes a unified description of the electromagnetic and weak interactions of fundamental particles.

In the process of β -decay, as studied by Hime and Jelley, the ^{35}S atom transmutes to a chlorine-35 ion ($^{35}\text{Cl}^+$), and emits both a neutrino (ν) and an electron (e^-). The neutrino interacts with matter so infrequently that Hime and Jelley's apparatus can catch no direct trace of it. Nevertheless, if the neutrino that escapes carries some mass with it, the law of energy conservation dictates that the energy of the remaining electron will be affected. From careful measurement of only the electrons' energies, which vary continuously from 0 to 167 keV (kiloelectron-

volts), Hime and Jelley have deduced that about 1 per cent of ^{35}S β -decays yield the heavy neutrino, and that the heavy neutrino has a mass of 17 keV.

Although small compared with the mass of the electron (511 keV) and proton (938 MeV), the mass of this neutrino is much larger than fashionable expectations, which tend to be near 10 eV. The evidence is a tiny kink in the electron energy spectrum near 150 keV (see figure). The remaining 99 per cent of β -decays in the experiment produce the neutrino associated with the electron, now known to possess a mass smaller than 9.4 eV. An experiment⁶ conducted six years ago at the Brookhaven National Laboratory implies that the neutrino associated with the muon could be emitted in at most 0.1 per cent of β -decays, and the conclusion follows that the 17 keV neutrino is the tau neutrino, if it is one of the fundamental trinity at all.

Fortunately, this identification of the 17 keV neutrino as the tau neutrino can be tested against astrophysical observations. There should be about 150 tau neutrinos in every cubic centimetre of space. These are relics, like the carefully measured background of microwaves⁷, from the Big Bang that created the Universe. They are so numerous that, if each has a mass of 17 keV, their combined mass would dwarf that of all the atoms in the Universe by at least a factor of 50. One effect of so much mass would be to brake the expansion of the Universe, but the astronomical data do not show the correct trend.

A let out would be if the primordial tau neutrinos somehow failed to survive; unfortunately, all known mechanisms for the decay of the neutrino provide a half-life about ten million times longer than the age of the Universe. Hypothetical particles and forces have been proposed to rid the Universe of the 17 keV neutrinos, but most proposals endow the neutrino with a very special property: that it be its own antiparticle. This possibility is unlikely, because the process of neutrinoless double β -decay (see Witherell's News and Views article⁸) would then proceed far more often than present data allow. On the other hand, if the 17 keV neutrino differs from its antiparticle, then supernova SN1987A would have emitted its burst of neutrinos in a shorter time interval than was measured by giant neutrino detectors here on Earth. Provocative as the astrophysical considerations are, they should not bias judgement of the 17 keV neutrino, and only evidence from well-controlled laboratory experiments need be weighed.

Abraham Pais has written⁹, "It is almost a rule that the first of a new generation of experiments on β -decay are wrong, even when performed by the best of physicists". The experiment of Hime and Jelley is the twelfth

1. Bates, H. W. *Trans. Linn. Soc. Lond.* **23**, 495-566 (1862).
2. Rothschild, M. *Nat. Hist. N. Y.* **76**, 44-51 (1967).
3. Turner, J. R. G. in *Dimensions of Darwinism* (ed. Grene, M.) 129-160 (Cambridge Univ. Press, 1983).
4. Ritland, D. R. & Brower, L. P. *Nature* **350**, 497-498 (1991).
5. Brown, K. S., Jr, Trigo, J. R., Francini, R. B., Barros de Moraes, A. B. & Motta, P. C. in *Plant-Animal Interactions* (eds Price, P. W., Lewinsohn, T. M., Fernandes, G. W. & Benson, W. W.) 375-402 (Wiley, New York, 1991).
6. Ackery, P. R. *Biol. J. Linn. Soc.* **33**, 95-203 (1988).
7. Brown, K. S., Jr *Revta Bras. Biol.* **44**, 435-460 (1985).
8. Davis, R. H. & Nahrstedt, A. in *Comparative Insect Physiology Biochemistry and Pharmacology* (eds Kerkut, G. A. & Gilbert, L. I.) Vol. 11 (*Pharmacology*) 635-654 (Pergamon, Oxford, 1985).
9. Raubenheimer, D. J. *chem. Ecol.* **15**, 2177-2189 (1989).
10. Platt, A. P. *Bull. ent. Soc.* **29**, 10-22 (1983).
11. Vane-Wright, R. I. *Biol. J. Linn. Soc.* **16**, 33-40 (1981).

with sensitivity to the 17 keV neutrino: the first² was performed by Simpson in 1985. Simpson implanted tritium (³H) atoms inside a detector made of silicon, and argued that the data showed a 17 keV neutrino emitted in about 3 per cent of the β -decays from tritium to helium, ${}^3\text{H} \rightarrow {}^3\text{He}^+ e^- \bar{\nu}$. Others pointed out that interactions between the e^- and the ${}^3\text{He}^+$ ion left in its wake mimicked the neutrino, and Simpson was obliged¹⁰ to reduce his estimate of 17 keV neutrino emission to 1 per cent. The total energy available from tritium, 18.6 keV, is too close to the mass of the proposed neutrino to permit a clear and unmistakable result.

Six of the ten experiments following Simpson's and preceding Hime and Jelley's exploited ³⁵S. The β -decay of ³⁵S releases 167 keV, sufficient to overwhelm the interaction between the outgoing electron and the ³⁵Cl⁺ ion left behind. Five saw no evidence of the 17 keV neutrino, but Simpson argues that four of these missed the particle because they distorted their data through use of a bad technique, namely, the measurement of electron momenta with a magnetic spectrometer. Although the electromagnetic distortions inherent in the β -decay process are not much smaller, spectrometer distortions must be evaluated from the data themselves. The fifth used a silicon detector¹¹; those data were later reinterpreted by Simpson who alleged that they showed evidence of the 17 keV neutrino. The sixth experiment, by Simpson and Hime, also used a silicon detector and saw evidence for the 17 keV neutrino¹².

The new experiment refines that arrangement, primarily by the insertion of masks between the ³⁵S source and the silicon detector. Electrons with kinetic energies smaller than 200 keV reflect in unpredictable directions from surfaces. The masks shield the silicon detector from stray electrons that follow errant paths from the radioactive source, and ensure that electrons enter the silicon detector straight on. Hime and Jelley surmise a statistical significance of the 17 keV neutrino of over eight standard deviations.

The failures to observe the 17 keV neutrino cast reasonable doubt on Hime and Jelley's conclusion. A null search by researchers at the Chalk River Nuclear Laboratories in Canada stands out as a painstaking effort¹³.

Their magnetic spectrometer measured energies nearly an order of magnitude better than the silicon of Hime and Jelley. In this experiment, as in most of those experiments that haven't observed the 17 keV neutrino, the radioactive source was mounted on a backing thicker than those of the 17 keV protagonists. Perhaps the signal got smeared out by backscattering. On the other hand, unusual peaks in the response of silicon detector systems to 624 keV electrons have been seen¹⁴, and attributed to scattering. Such a peak in the right place might deceive the protagonists.

The attribute of the positive experiment from Lawrence Berkeley³ is that no scattering occurs between the source and detector: carbon-14 (¹⁴C) has been grown right into the germanium crystal detector. The β -decay of ¹⁴C releases 156 keV, enough to avoid the problems encountered with tritium. The experiment has a modest data sample: the statistical significance of the 17 keV neutrino is only three standard deviations.

CELL BIOLOGY

Cell cycle gets more cyclins

Tim Hunt

ON page 512 of this issue¹, Motokura and colleagues describe the isolation of a complementary DNA of a putative oncogene, which appears to encode a polypeptide that is a member of the cell-cycle regulating cyclin family. The bearing of this discovery on oncogenesis is as yet uncertain, but, at the very least, those whose business it is to study control of the cell cycle have an important new lead to follow.

It is almost three years since the cyclin genes of clams and sea urchins were found to have counterparts in the *CLN* genes of budding yeast²⁻⁴. But whereas in higher eukaryotes and fission yeast the then-known cyclins controlled mitosis^{5,6}, in budding yeast they controlled Start, the decision point at the G1 $\rightarrow$ S transition of the cell cycle⁷. So, for three years budding yeast have had no mitotic cyclins and higher eukaryotes no Start (or G1) cyclins.

This unlikely and unsatisfactory state of affairs may now be over: in the 5 April issue of *Cell*, two groups^{8,9} describe the discovery of a set of mitotic cyclins in budding yeast; while Motokura and colleagues' peculiar-looking cyclin-like gene, PRAD1, which is over expressed in a rare form of benign parathyroid tumour, as a result of a chromosome rearrangement that placed this previously unknown gene under the control of a new promoter, is the best candidate so far for a human G1 cyclin. Whether PRAD1 is in any way responsible for the deregulated growth of the cells that comprise the tumour remains to be seen, but even if it is not, it enlarges the cyclin family in a most remarkable and intriguing way. Last year, a French group reported a reminiscent finding in a hepatic

The Rudjer Bošković Institute experiment⁴ has similar statistical significance, but uses a distinct physical process, the radiation that accompanies the capture of an atomic electron by germanium-71 and subsequent emission of a neutrino. This evidence contradicts a very similar earlier experiment¹⁵ performed by the same group, in which no evidence for the 17 keV neutrino was observed in the electron capture of iron-55.

The implications of a 17 keV neutrino are so important that final judgement should be withheld until all possible artefacts are empirically eliminated. This work is not yet complete. The words of C. D. Ellis and W. A. Wooster¹⁶, written in 1925 as they started the seminal experiment that led to the invention of the neutrino still apply. "Some very interesting phenomenon seems to be involved in β -ray disintegration..." □

Harry N. Nelson is in the Department of Physics, University of California, Santa Barbara, California 93106, USA.

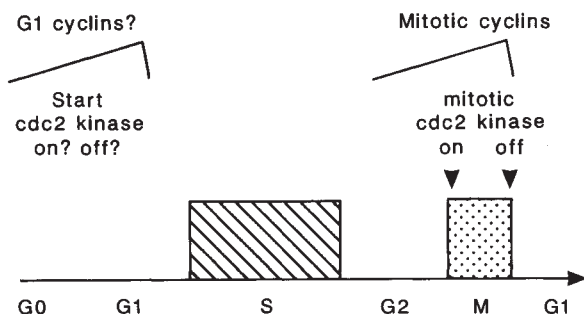
growth from a woman in whom a hepatitis B virus genome was inserted into her cyclin A gene so as to fuse a viral open reading frame with that of the cyclin¹⁰. In this case too, it is not known if this contributed to cellular transformation.

Cyclins were first identified by their property of sudden, cell-cycle regulated proteolysis, which fractionally preceded the onset of anaphase in the cleaving cells of clams and sea urchins. Later, it was discovered that they bound p34^{cdc2} and thus formed one of the subunits of maturation-promoting factor^{11,12}.

There are now over 30 cyclin sequences available for comparison, which fall into clear categories: A-type, B-type and the G1 cyclins. The B-type cyclins have been found in eukaryotic organisms from yeast to man. They associate with the p34^{cdc2} subunit encoded by the *cdc2* gene, and show maximal protein kinase activity during metaphase. Cyclin A is more mysterious. It seems to be a mitotic cyclin that accumulates and is activated and destroyed slightly earlier in the cell cycle than the B-type cyclins. It has not yet been found in yeast, and the only cyclin A deletion mutant, from *Drosophila*, is thought to arrest in G2 (ref. 13). Yet cyclin A has been implicated in S phase¹⁴, in cellular transformation¹⁰, and is able to complement G1 cyclins in budding yeast according to Steven Reed (personal communication).

This brings us to the difficult point. The current model for mitotic control can be summarized as follows (and see figure). Cells contain a variety of proteins whose properties alter when they are phosphorylated. When they are phosphorylated to a high

1. Hime, A. & Jelley, N. A. *Phys. Lett. B* **257**, 441-449 (1991).
2. Simpson, J. J. *Phys. Rev. Lett.* **54**, 1891-1893 (1985).
3. Sur, B. et al. Lawrence Berkeley Laboratory Preprint LBL30109.
4. Zlmen, I. et al. IRB Preprint 09 11/90 LBI.
5. Miller, D. J. *Nature* **349**, 374-387 (1991).
6. Ahrens, L. A. et al. *Phys. Rev. D* **31**, 2732-2736 (1985).
7. Lindley, D. *Nature* **343**, 207 (1990).
8. Witherell, M. S. *Nature* **331**, 392-393 (1991).
9. Pais, A. *Inward Bound* p. 11 (Clarendon Press, Oxford, 1986).
10. Simpson, J. J. & Hime, *Phys. Rev. D* **39**, 1825-1836 (1989).
11. Ohi, T. et al. *Phys. Lett.* **160B**, 322-324 (1985).
12. Hime, A. & Simpson, J. J. *Phys. Rev. D* **39**, 1837-1850 (1989).
13. Hetherington, D. W. et al. *Phys. Rev. C* **36**, 1504-1513 (1987).
14. Plansky, B. *Nucl. Instr. Meth.* **61**, 285-295 (1968).
15. Zlmen, I. et al. *Phys. Scripta* **38**, 539-543 (1988).
16. Ellis, C. D. & Wooster, W. A. *Proc. Camb. phil. Soc.* **22**, 849-860 (1925).



The cell cycle is represented as a time-line, with S phase and M phase represented as shaded boxes. The accumulation and destruction of mitotic cyclins is schematically represented as a graph, with destruction of the cyclin leading into interphase. The timing of accumulation and destruction of the G1 cyclins is largely hypothetical (see ref. 3).

enough degree, the cell enters mitosis. And when these same proteins are subsequently dephosphorylated, the cell leaves mitosis. The proteins remain in their mitotic state of phosphorylation for as long as the mitotic kinase(s) remain active. Simply put, high kinase = mitosis, low kinase = interphase. Cyclin is required, but is not sufficient, to activate the kinase activity of cdc2 for the G2 → M transition, and destruction of cyclin is what shuts off cdc2 kinase, causing the transition from mitosis to interphase.

Entry into S phase is not like this. Hartwell and colleagues showed that the budding yeast equivalent of cdc2, CDC28, is required for cells to enter S phase⁷. Cells with temperature-sensitive alleles of the gene responsible arrest at Start, implying that the activity of CDC28 kinase, although required for entry into S phase, is not needed to complete S phase once it is under way. Thus the mitotic model at first sight breaks down. Entry into S phase seems to be controlled in a different way. It is as though there is a period of the cell cycle of ill-defined length and unknown substance during which CDC28 has to be active. The period might be brief, almost a singularity in the cell cycle, or it might be long. Probably, CDC28 kinase, supported by its companion G1 cyclin-like subunits, has to stay active for a finite period, during which time progression of the cell cycle is latched by the synthesis of all kinds of components needed for DNA synthesis. Such a view is consistent with the well-known required period of exposure to growth factors — typically eight hours or so — for cells to perform the transition from resting to growing after a period of quiescence.

The question is, what are the key substrates of the G1 forms of CDC28 kinase? And does the proteolysis of G1 cyclins at any time form a signal in the same way as destruction of the B-type cyclins brings about exit from M phase? Or is it that rapid constitutive turnover of G1 cyclins keeps their levels low and thereby stops the Start-specific form of cdc2 kinase from exceeding the Start threshold?

It was the amino-acid sequences of CLN1,

2 and 3 that initially betrayed their resemblances to cyclins, not their cell-cycle behaviour. Yet the similarity to mitotic cyclins is faint. Of the roughly 400 residues in cyclins, only five widely but conservatively spaced amino acids show complete conservation in all 30 cyclin sequences from yeast to man. These residues probably mark important sites of contact with the cdc2/CDC28 subunits. Another important difference lies in the N termini of G1 cyclins, compared to mitotic cyclins of the A- and B-type family. Mitotic cyclins contain a 'destruction box' with a sequence such as RAALGNISN (single-letter code) that is typically located about 40 residues from the start of the protein and 80–100 residues upstream of the conserved 'cyclin box' (ref. 15). The G1 cyclins have no destruction box and have a shorter N terminus upstream of the cyclin box and a longer C-terminal extension beyond it. Their stability is thought to be determined by PEST sequences⁴.

So, where does PRAD1, the new cyclin-like gene from humans, fit into the scheme of things? It contains the five conserved residues, and shows ghostly sequence homology with cyclin A (the resemblance to CLN genes is poor). It clearly represents a new class of cyclin, however, if cyclin it be, *sensu stricto*. Like cyclin A, it can bind to and activate cdc2. Its transcript shows striking cell-cycle fluctuations, high in G2, low in S phase, but the protein levels have yet to be measured. If anything, these sparse data suggest the protein has a role in mitosis, but because the control of S phase — if indeed that is what we are talking about — is still so imperfectly understood, we may all yet be on the track of a boojum¹⁶.

Tim Hunt is at the ICRF Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK.

Tim Hunt is at the ICRF Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK.

1. Motokura, T. *et al.* *Nature* **350**, 512–515 (1991).
2. Hadwiger, J. A., Wittenberg, C., Richardson, H. E., de Barros Lopes, M. & Reed, S. I. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6255–6259 (1989).
3. Richardson, H. E., Wittenberg, C., Cross, F. & Reed, S. I. *Cell* **59**, 1127–1133 (1989).
4. Nash, R., Tokiwa, G., Anand, S., Erickson, K. & Futcher, A. B. *EMBO J.* **7**, 4335–4346 (1988).
5. Murray, A. W. & Kirschner, M. W. *Science* **246**, 614–621 (1989).
6. Nurse, P. *Nature* **344**, 503–508 (1990).
7. Hartwell, L. H., Culotti, J., Pringle, J. R. & Reid, B. J. *Science* **183**, 46–51 (1974).
8. Surana, U., Roberts, H., Price, C., Shuster, T. & Nasmyth, K. *Cell* **65**, 145–162 (1991).
9. Ghisla, J. B. *et al.* *Cell* **65**, 163–188 (1991).
10. Wang, J., Chenivresse, X., Henglein, B. & Bréchet, C. *Nature* **343**, 555–557 (1990).
11. Draetta, G. *et al.* *Cell* **56**, 829–838 (1989).
12. Gautier, J. *et al.* *Cell* **60**, 487–494 (1990).
13. Lehner, C. F. & O'Farrell, P. H. *Cell* **61**, 535–547 (1990).
14. D'Urso, G., Marraccino, R. L., Marshak, D. R. & Roberts, J. M. *Science* **250**, 786–791 (1990).
15. Glotzer, M., Murray, A. W. & Kirschner, M. W. *Nature* **349**, 132–138 (1991).
16. Carroll, L. *The Hunting of the Snark* (Macmillan, London, 1876).

Age and autophagy

Is it one of the trials of old age to be eaten from within by one's own proteases? Or is it even a cause? N. Schwartz-Ben Naim *et al.* (*Biochem. J.* **275**, 47–52, 1991) find that membranes of red blood cells from aged citizens contain proteolytic breakdown products of the main transmembrane protein, band 3. The effect can be reproduced *in vitro* by exposing the membranes to the endogenous protease, calpain I. The attack is on the large cytoplasmic domain of the band 3. Perhaps then the reduced life-span of red cells in old people is caused not by tired enzymes but rather by a change at the outer membrane surface, consequent on the proteolysis.

New for old

If the Universe collapses, might a new universe somehow be born? The inflationary model for cosmology offers apparently such a hope. Immediately after the Big Bang, inflation causes an exponential expansion, blowing up a tiny volume into the Universe observable now. In principle, when the Universe eventually contracts, the same microphysical processes could cause a time-reversed "deflation", abruptly shrinking a large volume of the cosmos to a tiny, dense region from which a new universe could emerge. Sadly, D. Goldwirth shows, in *Physics Letters* (**B256**, 354–358; 1991), that this mechanism for cosmic reincarnation is possible but highly unlikely. The virtue of inflation is that, starting from almost any initial conditions, it creates a smooth, uniform, expanding universe. For deflation, the opposite is true: it will start only if the initial conditions are precisely right.

Iron implications

T. A. Rouault *et al.* (*Cell* **64**, 881–883; 1991) draw attention to an intriguing similarity between an iron-regulated RNA-binding protein (IRE-BP) and the Krebs' cycle enzyme aconitase. IRE-BP binds to sequences in the messenger RNAs for the iron-storage protein ferritin and the transferrin receptor, in the first case inhibiting translation and in the second stabilizing the RNA. With excess iron the affinity for RNA is decreased, and more ferritin and less transferrin receptor is synthesized. Rouault *et al.* point out that the proteins' sequences coincide in regions which the known structure of aconitase implies are functionally important: aconitase is an iron-sulphur protein that can exist in an inactive state with only three iron atoms per molecule and an active state with four iron atoms per molecule. The authors say the proteins may be evolutionarily related and share a structural mechanism for iron-regulated activity.

A smooth ride for galaxies

M. G. Edmunds & S. Phillipps

SEARCHES for large-scale distributions of matter in the Universe are all the rage nowadays, as Dressler indicated in a review article in last week's issue¹. But the picture of how these distributions are gravitationally pulling our Galaxy and its neighbours around our corner of the Universe seems to be much simpler than has been supposed, calculations by James, Joseph and Collins suggest². The difficulties seem to have arisen from biases inherent in the sampling necessary to make large-scale surveys manageable.

The expansion of the Universe has been known since Edwin Hubble's fundamental discovery of a regular relationship between how far away a galaxy is, and its velocity away from us. Except for the very nearest galaxies, the velocities increase almost linearly with distance. The "almost" qualification comes from the slight inevitable variation of individual galaxy velocities about the mean expansion, just as one would not expect all atoms in an expanding gas to have exactly the same speed. The basic origin of a galaxy's extra motion is the perturbing effect of the gravitational pull of the galaxies or clusters of galaxies in its neighbourhood. In recent years such motions have been intensively studied.

When considering the motion of individual galaxies relative to the general expansion, there is one fundamental reference frame — the microwave background radiation left over from the early Universe, predating the development of significant deviations from a Hubble flow. The Doppler shift of the microwave background (essentially a slight temperature anisotropy) shows that our little group of galaxies (which is a few megaparsecs in size) is moving at about 600 km s^{-1} relative to this 'fixed' frame in a direction which, viewed from Earth, is towards the constellation Hydra.

Some years ago, before our motion relative to the microwave background had been determined, a claim was made³ that our local group showed a motion of some 450 km s^{-1} relative to moderately distant galaxies ($35h^{-1}$ to $65h^{-1}$ Mpc away, where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and probably $0.5 \leq h \leq 1$). This apparent motion, known as the Rubin-Ford effect, turned out not to be in the same direction as our motion through the microwave background. Taken literally, this would imply a gross streaming motion — relative to the microwave background — of the galaxies in a very large local volume of space, some $100h^{-1}$ Mpc in size. The cause of such large-scale streaming was unknown. Although it stimulated a lot of useful work, it now seems that the original Rubin-Ford effect should not be interpreted as implying a streaming motion at all. As so often in observational astronomy, it appears that the selection of the particular sample biased the results.

In a rather convincing paper, James, Joseph and Collins² argue that the basic problem came from taking only galaxies with a restricted range of apparent brightness (as seen from Earth) and in using their exact brightness to estimate the distances of the galaxies, on the assumption that all of the particular type of galaxy used have the same intrinsic brightness. The trouble with this procedure is that real galaxies do have a distribution in intrinsic brightnesses, and the result — as discovered in surveys of stars in the 1920s by the Swedish astronomer Malmquist — is that the distant galaxies observed in the survey tend to be intrinsically brighter than average, and the nearest ones are on average intrinsically fainter. If the distances are being deduced from these brightnesses, then the redshifts (the measure of velocity) of the nearer galaxies seem anomalously low and the redshifts of the furthest ones seem anomalously high — and the difference is attributed to a systematic motion. Combine this with differences in average distances to galaxies in different directions in the sky, attributable to the way the sample was chosen or to galaxy clustering, and an apparent flow in some direction can result.

What James, Joseph and Collins have done is to ask the question: if the sample of galaxies is really moving smoothly with the Hubble flow, and the galaxies have a reasonable intrinsic spread of brightnesses, and the galaxies are chosen as they were — on the basis of their velocities relative to us — then what sort of velocity distribution would we expect to see? From simulations they conclude that, even with smooth Hubble flow, the sample would appear to have spurious streaming flows of around $600 - 800 \text{ km s}^{-1}$. This pleasing result brings us back to the 'simplest' picture of regular Hubble flow without additional large-scale velocity fields.

But what about the 600 km s^{-1} motion of our local group relative to the microwave background? The interpretation may have become considerably clearer following a survey of redshifts⁴ based on a catalogue of galaxies detected as infrared sources by the IRAS satellite. The claimed advantage of using this sample is that it is less prone to the systematic selection effects which can plague optical samples. Analysis of the velocities in the sample, out to about $200h^{-1}$ Mpc, gives a picture of general expansion, superimposed on which are local perturbations due to the gravitational effect of 12 massive clusters or superclusters of galaxies. The 600 km s^{-1} of our local group is mainly due to the attracting effect of the great clusters in the direction of Virgo, Hydra-Centaurus, Pavo and Fornax-Eridanus, and whose centres are within $65h^{-1}$ Mpc of us.

The only problem is the implied extraordinary mass and size of the clusters. They must

contain a great deal of dark matter (perhaps 30 times that visible as galaxies) and extend out $30h^{-1}$ Mpc from their centres — which may be compared with their 'visible' size, a few (perhaps $3h^{-1}$) megaparsec. At this size, many of the big clusters essentially merge into each other. Perhaps this should not be regarded as a problem — after all, it was difficulties with the high velocity dispersions in clusters that first led Zwicky to identify the problem of 'missing mass' in the 1930s.

It now seems possible to dispense with the idea that there might be some additional source of perturbing gravitation — the so-called Hidden Great Attractor — pulling on the galaxies in our neighbourhood, the observed clusters themselves being massive enough to account fully for the local group motion. The samples⁵ which gave rise to the idea of a Great Attractor may also suffer from selection effects, in particular they used elliptical galaxies which biased the samples towards galaxy cluster members, as ellipticals are far more common in clusters than in the general field between clusters.

It also seems⁶ that the method of estimating distances to the ellipticals (crucial when their Hubble flow velocities are estimated from their distances) may involve large random and systematic errors. The distance estimator uses the fact that the (fairly easily measured) internal velocity dispersion of the stars in a galaxy is directly related to its intrinsic brightness: both essentially depend on the mass of the galaxy. But the relation seems to vary between different clusters, perhaps because of the different histories or environments of the galaxies in particular clusters, and hence cannot be used as a reliable universal indicator of distance.

Does this mean that there is no Great Attractor at all? Semantics seems to have confused the issue. The "group of seven"^{1,5} still support the concept, but identify it with the mass associated with the galaxies in that part of the sky. The argument now seems to be about whether the local flow is said to be due to the mass in a particularly dense concentration, or to the combined effect of several large clusters.

The kinematics of galaxies — and of galaxy clusters — in the Universe may soon be presenting a rather more straightforward picture of expansion than was threatened a few years ago. What remains is the problem of the size and the nature of the mass in the great clusters, although the very fact of the concentration of the mass into these structures must surely provide a valuable clue. □

M. G. Edmunds and S. Phillipps are in the Department of Physics, University of Wales, PO Box 913, Cardiff CF1 3TH, UK.

1. Dressler, A. *Nature* **350**, 391-397 (1991).
2. James, P. A., Joseph, R. D. & Collins, C. A. *Mon. Not. R. Astr. Soc.* **248**, 444-450 (1991).
3. Rubin V. C., Ford W. K., Thonnard, N., Roberts, M. S. & Graham, J. A. *Astron. J.* **81**, 687-697 (1976).
4. Rowan-Robinson, M. et al. *Mon. Not. R. Astr. Soc.* **247**, 1-19 (1990).
5. Lynden-Bell, D. et al. *Astrophys. J.* **326**, 19-49 (1988).
6. Lucey, J. R., Gray, P. M., Carter, D. & Terlevich, R. J. *Mon. Not. R. Astr. Soc.* **248**, 804-815 (1991).

Supercurrents at a pinch

John Gallop

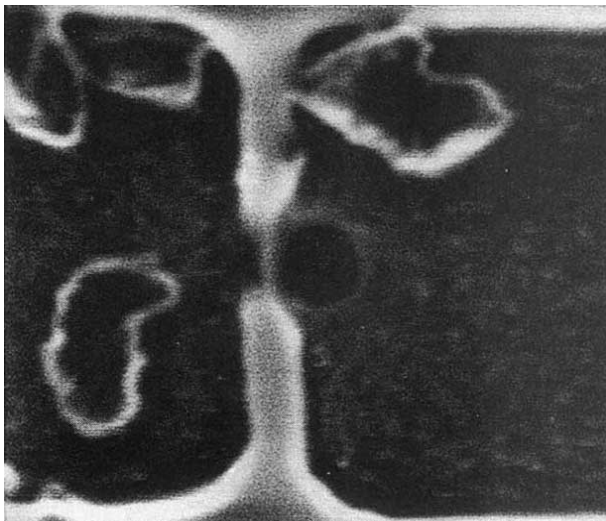
IF ANY one property is expected of a superconductor it is that it should carry high currents without resistive loss. In the case of the high-temperature superconductors the promise of their dramatically increased transition temperatures (up to 125 K, compared with 23 K for the best conventional superconductors) was initially tempered by the experimental observation of rather low critical currents, that is the maximum value passed before dissipation becomes detectable. But in *Physical Review Letters* last week (66, 1785–1788; 1991), W. Jiang *et al.* amply confirmed what many workers in the field had come to believe: that it will not be the intrinsic current carrying capabilities of the high-temperature materials that will prevent their widespread technological application. The authors report that reducing the width of the current-carrying channel dramatically increases a superconductor's current-carrying capability, as characterized by the critical current density.

Jiang *et al.* set out to see the effect of passing a supercurrent through a narrow constriction in thin films of $\text{YBa}_2\text{Cu}_3\text{O}_7$, the ubiquitous high-temperature superconductor with a transition temperature T_c of 91 K. First, a coarse pattern was chemically etched in films 200–500 nm thick, before a finely focused beam of gallium ions was used to mill narrow bridges 50–400 nm across (see micrograph). The authors found that reducing the width of the bridge increased the critical current density the material could carry. Even at a width of 50 nm, the critical current j_c had not reached a plateau, contrary to expectations, causing the authors to put forward a new model for the current limit.

Several mechanisms are known to limit the critical current of a superconducting filament. Perhaps the most fundamental is the critical 'depairing' current density: superconductivity occurs when electrons 'condense' into loosely bound pairs, and the depairing j_c can be understood simply as that value of current for which the local kinetic energy density of the pairs becomes greater than the condensation energy density. The latter is simply related to the thermodynamic critical magnetic field H_c , which also suppresses superconductivity. The relatively low values of the depairing j_c found in elemental superconductors such as tin or indium agree approximately with this simple model which links the low-critical fields for such materials with their limited ability to carry current.

In some so-called type II superconductors, to which category belong all of the high- T_c cuprates so far investigated, there is another

critical magnetic field, lower than the thermodynamic one. Application of a field greater than this does not destroy superconductivity, but causes discrete quantized lines of magnetic flux to enter the material. The microscopic structure of the resulting fluxons can be described by two characteristic lengths which are material dependent. In a thin core of material of radius ξ (the coherence length) the superconductivity is



A micrograph of a 50-nm bridge (centre) milled into thin-film $\text{YBa}_2\text{Cu}_3\text{O}_7$, giving superior current-carrying characteristics. (Courtesy, A. Widom.)

quenched and the quantized fluxon threads this region. This core is surrounded by a region of circulating supercurrent which extends over an annular length λ , the penetration depth of the superconductor.

The ability of a type II superconductor to remain lossless in the regions between the normal fluxon cores is necessary but not sufficient for high-current or high-field applications. One extra trick is needed: to fix the lattice of fluxons. In the presence of a current through the superconductor these miniature magnets experience a force at right angles to both the direction of current and to the direction of the flux quantum (as a result of Ampere's law). Motion of the fluxons produces a voltage by induction along the superconductor, with resulting dissipation. In an ideal material this flux flow would occur at arbitrarily low current densities. In real materials, fluxons are pinned by defects and impurities, with no movement allowed until the driving force, proportional to current density, can overcome the pinning forces on fluxons. The critical current density arising from this source of dissipation is a function of the strength of the pinning and also of temperature.

To achieve practical high current densities in bulk samples of superconductor the main thrust has been to introduce inhomogeneities in the crystal lattice which are effective in

pinning flux. But it has been recognized for many years that another approach, not obviously of practical use, may also be followed. By making the cross-sectional area of the conductor smaller one may exploit an energy barrier which grows to prevent fluxons entering the conducting channel, and this is the approach taken by Jiang *et al.* The exclusion of fluxons from the bridge then removes the restriction on critical current due to fluxon flow, thereby allowing the depairing limit to be approached.

The characteristic lengths λ and ξ , which determine the depairing limit, are somewhat sample dependent and difficult to measure; more seriously, they are highly anisotropic in the layered cuprate superconductors. The highest value of j_c that Jiang *et al.* measure, $1.3 \times 10^9 \text{ A cm}^{-2}$, is a factor of 10 higher than has been previously reported for any other superconductor, whether conventional or cuprate; but it is still less than the depairing limit calculated using reasonable values of the characteristic lengths (with $\lambda = 100 \text{ nm}$ and $\xi = 1 \text{ nm}$, one expects $j_c = 3 \times 10^{10} \text{ A cm}^{-2}$).

Because j_c continued to increase even when the width of the bridge had been reduced to 50 nm, Jiang *et al.* propose that another dissipation mode is operating. This involves the generation of fluxons in a fashion analogous to the process that creates vortices in superfluid helium, the Onsager–Feynman process. (R. P. Feynman *Prog. Low-Temp. Phys.* 1, 17–53; 1955).

Although the presence of fluxons in a small bridge is energetically forbidden, the current density may be sufficiently high to generate fluxon rings (or more likely, in the highly anisotropic cuprate materials, a pair of oppositely directed fluxons) around the exit of the bridge, where it joins the electrode, rather as a smoker blows smoke rings. The critical current density determined this way depends on the usual two characteristic lengths, but it also varies with the width of the bridge.

Does this result have practical implications? The short, ultrafine superconducting tracks produced in the experiments seem to have little in common with the vast lengths of rugged, unbroken, high-quality conductor which would be required for magnets or power transmission. However, one interesting potential use of the new materials, and one of the most demanding in terms of the critical current density required, is to provide interconnects in very fast cryogenically cooled chips. The demonstrated ability to produce a 50-nm-wide track capable of carrying a current of 100 mA without dissipation might prove extremely useful in this context. □

John Gallop is at the National Physical Laboratory, Queens Road, Teddington TW11 0LW, UK.

Naming names in botany

Clive A. Stace

THOSE who have been frustrated by a change in the Latin name of an organism can take heart from events at a recent meeting* — determined efforts are in hand to solve the problem, despite ample evidence that there are still wide differences of opinion on the best way to achieve it.

There was, however, no disagreement that names are important: "the name of a plant is the key to its literature" is as apposite today as when Van Steenis proclaimed it in 1957. If one uses the wrong name, or if different names are in use at once, confusion is the result. For example, *Lycopersicon esculentum*, *L. lycopersicum* and *Solanum lycopersicum* are all names used for the tomato. Although employed for a short period recently, the second is a nomenclatural synonym of the first and should not be used, but the first and third are equally correct names according to the genus into which the tomato is placed.

The subject matter of the meeting was very much plant (and fungal) names, and the few contributions by zoologists and bacteriologists added little to the proceedings. Animal and microbial nomenclatural problems are broadly the same as plant ones, but the backgrounds, and so the solutions, differ in detail. More bizarre suggestions that numbers should be used instead of names, or Esperanto instead of Latin, are unlikely to be followed up.

A careful distinction has to be drawn between nomenclatural and taxonomic name changes. Nomenclatural name changes are the result of application of the *International Code of Botanical Nomenclature*, especially the rules of priority; provided the *Code* has been interpreted properly nomenclatural changes are mandatory. Taxonomic changes occur as a result of new research, or new assessments of data, and are often open to more than one interpretation. They are therefore always to a certain degree subjective and often the evidence for them is equivocal. Nonetheless there was general agreement at the meeting that it would be foolish to legislate against taxonomic changes. To do so would ossify taxonomic data and make further research almost pointless — it would make as much sense as deeming a particular metabolic pathway, or a certain DNA sequence, correct in detail, even after further study had shown that it was not. The best way to prevent the ambiguity caused to users by taxonomic name changes would be for specialist user groups (horticulturists, timber technologists, weed researchers, conservation legislators and so on) to issue check-lists of standard names of their relevant species. Such a list could last for an agreed number of years until it was rendered too outdated by

the changes that had accumulated; a new list could then be drawn up.

But what of ways of combating unnecessary nomenclatural name changes, which may be the result of meticulous rummaging through the old, neglected literature, or indeed reversion to a formerly used name, which, after all, is again deemed correct? Two schools emerged. First, the radicals (such as R. K. Brummitt, Kew; C. R. Gunn and J. H. Wiersma, US Department of Agriculture; W. Greuter, Botanical Garden and Museum Berlin; V. H. Heywood, International Union for the Conservation of Nature; and myself) who would opt for a list of accepted or protected names that could not be predated, nor whose spelling or gender could be changed by bibliographic delving. Second the reactionaries (exemplified by W. R. Anderson, Univ. Michigan; A. Cronquist, New York Botanical Garden; F. R. Fosberg, Smithsonian; and D. J. Mabberley, Univ. Oxford), who felt that the problems are exaggerated and are best dealt with by rigorous application of the *Code*.

There seemed to be a clear majority who were not persuaded by this second view — promises for over half a century that close adherence to the *Code* would bring stability have not been fulfilled, and changes to the *Code* (described by several speakers as ill-conceived meddling) designed to remove anomalies have often caused yet more instability. But had the meeting taken place on the other side of the Atlantic, the majority might well have supported the opposite opinion.

The main meeting was followed by closed sessions of specialist committees, whose task is to hammer out proposals to be put to the nomenclatural sessions of the next International Botanical Congress (Tokyo, 1993). Whatever the difficulties caused by nomenclatural instability, and whatever the prevailing view among the global taxonomic community, proposals to adopt lists of protected names will have to be passed by voting at Tokyo. There is at least an even chance that they will not.

Meanwhile, as reported last year (*Nature* 348, 275; 1990), another, larger group of taxonomists is about to launch an attempt to produce a world check-list of higher plant species in five–six years. This scheme goes under the provisional title of the Species Plantarum Project. If it is successful I believe that the pressure to adopt it as a list of protected names will become irresistible, and that this route offers the best hope for greater stability of plant names — and of giving taxonomists a better name among their fellow biologists. □

Clive A. Stace is in the Department of Botany, University of Leicester, Leicester LE1 7RH, UK.

Signs of stress

STRESS is the great curse of modern life. It is regularly blamed for migraine, professional burn-out, depression, premature ageing, and much else besides. Daedalus now intends to pin down and quantify this vague but almost universal malady.

In conventional engineering stress-analysis, a transparent model of a structure or component is put under load. At each point in the model, the local stress-induced anisotropy rotates the plane of polarization of the light traversing it. Viewed through a polarizer, the model shows its stress distribution as a beautiful fringe pattern.

Mental stress, says Daedalus, also leaves its physical mark on the sufferer's body. Habitually hunched shoulders, knotted stomach, clenched fists, high blood-pressure and so on, are obvious examples. Milder chronic stress may simply tense up a few specific internal muscles somewhat beyond their normal 'tone'. Human beings are practically opaque to polarized light. But they are transparent to polarized ultrasound, which can also travel through solids and be plane-rotated by stress-anisotropy.

So Daedalus's ultrasonic body-stress mapper places a transversely polarized ultrasonic source on the patient's chest, say, and examines the resulting ultrasonic shadow-graph on his back. To 'develop' the pattern, Daedalus has a special liquid-crystal 'stress-varnish'. Nematic liquid crystals align themselves with the microscopic 'grain' of a surface, and if painted on the skin should sensitively reveal the direction of vibration at each point. Through polarized sunglasses the watching psychiatrist will see the internal web of muscles and tendons outlined on the patient's back, darkening or lightening as they tighten or relax.

Experience will be needed to interpret such maps. Bottlers-up of suppressed rage, or fear, or guilt, or anxiety, or despair, or hopeless longing, will all show characteristic internal stress-maps, obvious once they have been identified. Facial stresses, revealed by an ultrasonic exciter on the back of the head, will be particularly revealing. DREADCO's liquid-crystal cosmetics will display the true tensions behind the fixed careless smiles or carefully neutral expressions so many of us maintain.

Even better, these stress-maps will change or intensify under verbal probing. The psychiatrist will rapidly home in on the source of all the trouble. He will even be able to tell when his wise counsels or deep insights are having some affect. But probably the best cure for stress would be to let the patient see his own stress-map in a mirror. He could then discover what thoughts or situations wind him up, and learn to counter them by active or meditative bio-feedback.

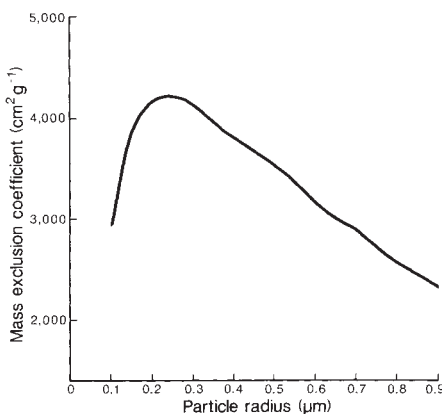
David Jones

* Improving the Stability of Names: Needs and Options Royal Botanic Gardens, Kew, UK, 20–22 February 1991.

Ice particles and the greenhouse

SIR — Within the framework of a suitable model the properties of the Earth's greenhouse can be calculated from tables of absorptivities of the greenhouse gases¹. We find that, taking the Earth's albedo to be 0.30, the Earth's effective infrared temperature with no greenhouse would be about 255 K. Adding the present-day greenhouse gases raises the Earth's mean temperature to about 292 K, in agreement with experience. Doubling the present atmospheric CO₂ content would raise the mean temperature by a further 1.5 °C without including a knock-on-effect from water vapour. The latter is what one cares to make it. A typical increase of a factor of 2 gives a rise of about 3 °C.

Even slight variations of circumstances that have been kept constant could alter the result of this investigation; the albedo is one such factor. Increasing the Earth's albedo from 0.30 to 0.315 would conceal the doubling of CO₂. The albedo depends on clouds in the troposphere, and on the nature



The backscatter mass exclusion coefficient of spherical ice particles averaged over the solar spectrum and the sunlit hemisphere, as a function of particle radius².

of the Earth's surface — oceans, deserts, grasslands, forests, crops, ice and snow.

Small ice crystals in the mesosphere also significantly affect the albedo. The figure shows the average mass coefficient for the exclusion of solar radiation due to backscattering by spherical ice grains uniformly distributed over a sunlit hemisphere². The computations are based on Mie theory combined with integrations over a hemisphere and an averaging with respect to the solar spectrum. We note that for particles with radii of a few tenths of a micrometre, a typical value of the mass exclusion coefficient is about 4,000 cm² g⁻¹, the result being insensitive to the precise particle size. Taking the mass loading of the particles over the whole Earth to be 10¹² ξ g (where ξ is a parameter that ranges from about 0.1 to 10), the albedo is increased by 10¹² ξ × (4,000) / 5 × 10¹⁸ = 8 × 10⁻⁴ ξ (the surface area of the Earth being ~ 5 × 10¹⁸ cm²). A value of ξ of order 10 would therefore yield a cooling of the Earth sufficient to cancel out a doubling of the atmospheric CO₂

content. Although such a value of ξ is not currently operative, the total water vapour content of the high atmosphere would, if condensed, be far more than sufficient. The measured mid-stratospheric water vapour content is indeed in excess of 10¹⁵ g. Conditions for the condensation of water vapour into particles appear to be delicate³, suggesting that no great change would be needed to increase ξ substantially. Episodes of extensive ice-crystal condensation could be triggered by either terrestrial or extra-terrestrial

African fishes

SIR — Meyer *et al.* conclude¹ that morphological diversification in flocks of cichlid fishes in Lake Victoria occurs without much molecular evolution. This agrees with our own data on hybrids between different species of cichlid fish from Lake Victoria, which can be bred easily over several generations²⁻⁴. In addition to the molecular data on mitochondrial DNA, our results show the genetic compatibility for the entire genome of some Lake Victoria 'haplochromines'.

But we disagree with Meyer *et al.*'s second conclusion that Lake Malawi species are more closely related to those of Lake Victoria than to those of Lake Tanganyika. We tested by hybridization the genetic compatibility of *Astatotilapia burtoni*, a riverine species from the Lake Tanganyika basin, with *Astatotilapia nubilus*, a riverine species from the Lake Victoria basin. The fertility of the hybrids between these species suggests that they are genetically almost as compatible as haplochromines from the Lake Victoria basin^{2,3}. Moreover, we also tried to breed hybrids between the monotypic genus *Astatoreochromis alluaudi* and several haplochromines of Lake Victoria. According to Meyer *et al.*, this species should be more closely related to Lake Victoria haplochromines than to those of Lake Tanganyika. We succeeded, however, in breeding only one hybrid population between *Astatoreochromis alluaudi* and *Astatotilapia 'black lividus'* (a so far undescribed species of Lake Victoria). These hybrids reproduced, but the eggs deteriorated before hatching. This is the only evidence for sterile hybrids that we have obtained so far. *Astatoreochromis alluaudi* is also the only species in Lake Victoria that stands out in its protein characteristics⁵.

We believe that Meyer *et al.* should include molecular data from the riverine Lake Tanganyika haplochromine *A. burtoni* in their analysis. Their second conclusion will then be that some haplochromines of different lakes arose from more closely related riverine species; in other words, that segregation had started before the lakes formed. The major problem of the species flocks of the great African lakes is clearly how they have evolved and are maintained in sympatry,

factors connected with the supply of condensation nuclei. Perhaps this is what happened during the Little Ice Age.

F. HOYLE

N. C. WICKRAMASINGHE

School of Mathematics,
University of Wales College
of Cardiff,
Cardiff CF2 4AG, UK

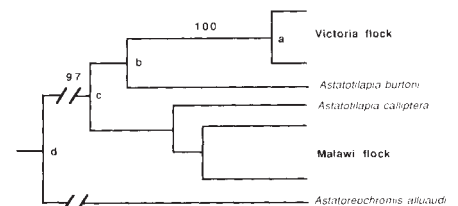
- Howard, J. N., Garing, J. S. & Walker, R. G. in *Handbook of Geophysics and Space Environments* (ed. Valley, S. L.) (McGraw Hill, 1965).
- Hoyle, F. & Wickramasinghe, N. C. *Earth Moon and Planets* (in the press).
- Gadsden, M. J. *Atmos. terr. Phys.* **52**, 247 (1990).

while obviously relying exclusively on pre-mating barriers — a significant ethological problem in its own right⁴.

M. — DOMINIQUE CRAPON DE CAPRONA
BERND FRITZSCH

Creighton University,
Department of Biomedical Sciences,
Omaha, Nebraska 68178, USA

MEYER *ET AL.* REPLY — Inferring relationships from patterns of hybridization is risky. The hybridization studies involving *Astatotilapia burtoni* males and *Astatotilapia nubilus* females produced only lethal hybrids². Females of *A. burtoni* mated rarely with males of *A. nubilus*, but produced viable hybrids². Numerous studies have shown that hybrid incompatibility, although usually correlated



Evolutionary tree based on comparisons of part of the control region of mtDNA. The region sequenced and the methods used are described by Meyer *et al.*¹. Bootstrap values appear above the two internal branches that are statistically significant (a—b and c—d). The distance from any node to the tips of the descendant branches are roughly proportional to the average number of base substitutional differences between species united by that node. The root of the tree (node d) was determined by using *Julidochromis* as an outgroup (see ref. 1). The two sequences of *Astatotilapia* have been deposited in Gen Bank with accession numbers X58151 (A.c.) and X58152 (A.b.).

with phylogenetic distance, is not an accurate metric for building trees and time scales.

In anticipation of questions about *Astatotilapia*'s phylogenetic status, one of us (A. M.) sequenced the most variable part of the control region of mitochondrial DNA from two key species, *A. burtoni* and *A. calliptera* (a non-endemic cichlid from Lake Malawi). Three *Astatotilapia* species from the Lake Victoria basin (*A. nubilus*, *A. elegans* and *A. piceatus*) had already been tested¹. The figure shows a tree relating the two new sequences to those of the species flocks in

Lakes Victoria and Malawi as well as to *Astatoreochromis alluaudi*. The results confirm (with 97 per cent confidence) the hypothesis, suggested by interspecific hybridization studies, that *A. burtoni* is a closer relative of these flocks than is *Astatoreochromis*. This analysis also suggests (although not significant statistically) that *A. burtoni* is closer phylogenetically to the Lake Victoria flock than is the Lake Malawi flock.

Furthermore, and contrary to the hypothesis of Crapon de Caprona and Fritzsche above, the members of the Lake Victoria flock (including *A. nubilis*, *A. elegans* and *A. piceatus* are more closely related to one another mitochondrially than to *A. burtoni*. The bootstrap value that shows the reality of the a-b lineage in the figure is 100 per cent. We infer that the Lake Victoria flock arose long after the splitting of the *A. burtoni* lineage from the a-b lineage. Our molecular estimate of the time of most recent common ancestry for the Lake Victoria flock is 200,000 years¹, which implies that the flock is younger than the lake (250,000 to 750,000 years old⁶). Thus, our results cast doubt on Crapon de Caprona and Fritzsche's model, above, according to which segregation occurred within the flock before the lake formed.

Another notable feature of the figure is

'Life' not critical?

SIR — Bak *et al.*¹ have presented evidence that the cellular automaton, the 'game of life', develops into a self-organized critical state², characterized by a $D(T) \sim T^{-1.6}$ distribution of times T required for the lattice to return to equilibrium following a random single-site perturbation. This power law implies an infinite expected equilibration time $\langle T \rangle$. But we believe that this behaviour is an artefact resulting from the relatively small lattices (100 by 100) used in ref. 1.

Self-organized criticality describes situations in which a spatially extended dissipative system spontaneously adjusts itself, without deliberate tuning of any external parameter, into a state with no finite correlation length or relaxation time, like an equilibrium statistical-mechanical system at its critical point. In a classic example², the addition of sand grains one at a time to an idealized sandpile increases the slope until a critical slope is reached, after which the arrival of an additional grain is likely to trigger an avalanche of any size, up to the size of the entire system, maintaining the mean slope thereafter at the critical value. A general mechanism of such criticality has been elucidated for models having a conserved quantity, such as the amount of sand. But Bak *et al.* suggest¹ that systems that lack any evident conserved quantity may also be critical.

On a finite $L \times L$ lattice, large avalanches are not possible, and the expected equilibration time $\langle T \rangle$ implied by the $T^{-1.6}$ power law is no longer infinite but should increase with L at least as $L^{0.4}$, because the initial per-

turbation can increase in diameter at most linearly with time. Our studies on lattices up to $1,024 \times 1,024$ with periodic boundary conditions (Fig. 1a) show instead that $\langle T \rangle$ approaches a constant value 200 ± 10 for lattices larger than about 100×100 . This suggests a characteristic extinction length of ~ 50 lattice points for perturbations propagating outward from their point of origin.

AXEL MEYER
Department of Ecology and Evolution,
State University of New York,
Stony Brook, NY 11794, USA

THOMAS D. KOCHER
Department of Zoology,
University of New Hampshire,
Durham, NH 03824, USA

ALLAN C. WILSON
Division of Biochemistry and
Molecular Biology
University of California,
Berkeley, CA 94720, USA

1. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. *Nature* **347**, 550–553 (1990).
2. Crapon de Caprona, M.-D. & Fritzsche, B. *Neth. J. Zool.* **34**, 503–538 (1984).
3. Crapon de Caprona, M.-D. *Ann. Sc. Zool.* **251**, 117–120 (1986a).
4. Crapon de Caprona, M.-D. *J. Fish. Biol.* **29**, 151–158 (1986b).
5. Verheyen, E., Van der Linden, A. & Decler, W. *Ann. Sc. Zool.* **257**, 93 (1989).
6. Temple, P. H. *Biol. J. Linn. Soc.* **1**, 363–371 (1969).

turbation can increase in diameter at most linearly with time. Our studies on lattices up to $1,024 \times 1,024$ with periodic boundary conditions (Fig. 1a) show instead that $\langle T \rangle$ approaches a constant value 200 ± 10 for lattices larger than about 100×100 . This suggests a characteristic extinction length of ~ 50 lattice points for perturbations propagating outward from their point of origin.

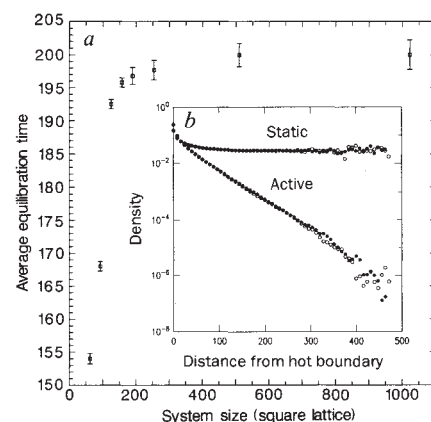


Fig. 1a, Average decay time for point perturbations in the game of life, as a function of linear system size, for square arrays with periodic boundary conditions. Perturbations decaying in ≤ 6 steps are excluded from the average. b, Semi-logarithmic plot of static and active density as a function of distance from the hot boundary, in a 488×256 array with a hot boundary on the left, a free boundary on the right and periodic boundary conditions in the vertical direction. Dots and circles show results of two independent runs of about 30 million steps each.

To determine this extinction length more precisely, we measured active and static densities (see below) in other runs (Fig. 1b) as a function of distance from an inhomogeneous 'hot' boundary, in which the configuration was randomized every 16 steps. The static density approached a characteristic equilibrium value of 0.03 away from the boundary, whereas the active density (here defined as live sites that were not alive 6 steps previously) declined exponentially with a scale length of 42 ± 3 , indicating a uniform rate of extinction per unit distance for perturbations propagating away from the hot boundary.

Our results indicate that the game of life is subcritical, with a large but finite relaxation time of 200 ± 10 steps and an extinction length of 42 ± 3 lattice points, and that the equilibrium state, restored after the decay of perturbations, is a roughly spatially uniform distribution of decoupled local oscillators with a density of ~ 0.03 . The possibility remains, however, that this apparent equilibrium may be metastable with respect to nucleation events too rare to have occurred in any of our simulations so far. This would be the case, for example, if the game of life allowed the existence of configurations analogous to Belousov-Zhabotinsky spiral cores, which cannot be destroyed from outside but instead entrain their surroundings in coherent waves of minimum period⁴.

CHARLES BENNETT
IBM T. J. Watson Research Center,
Yorktown Heights,
New York 10598, USA

MARC S. BOURZUTSCHKY
Department of Physics,
California Institute of Technology,
Pasadena, California 91125, USA

1. Bak, P., Chen, K. & Creutz, M. *Nature* **342**, 780–782 (1989).
2. Bak, P., Tang, C. & Wiesenfeld, K. *Phys. Rev. A* **38**, 364–381 (1988).
3. Hwa, T. & Kardar, M. *Phys. Rev. Lett.* **62**, 1813–1816 (1989).
4. Fisch, R., Gravner, J. & Griffeath, D. *Threshold-Range Scaling of Excitable Cellular Automata* (University of Wisconsin preprint).

Scientific Correspondence

SCIENTIFIC Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words and five citations, and to contributions using simple language.

If new results are being described, priority is generally given to communications that do not describe work in which the author is involved. Authors of contributions of this type should explain in a covering letter why theirs has a particular claim on *Nature's* space. Contributions may be sent to referees and, in the case of matters arising from material published in *Nature*, are sent to the author of that article for comment. A more detailed guide to authors is available from Washington or London. □

Formation of cosmic structure by Doppler instability

Craig J. Hogan*

Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606, Japan

A new mechanism is described which can create an instability in homogeneous gaseous matter at very low density. When an isotropic background radiation field has, near an electronic resonance (such as the hydrogen Lyman- α line), a spectral feature for which photon occupation number increases with frequency, moving atoms increase their speed by taking energy from the photon distribution. In a cosmological setting, a sufficiently intense spectral feature can interact with neutral atomic gas, after recombination, to generate protogalactic perturbations of the scale and magnitude needed to explain large-scale cosmic structure.

THE process of gravitational instability in an expanding universe—the growth of small-amplitude perturbations and their collapse into bound astronomical systems—accounts in broad terms for the formation of galaxies and many of the statistical properties of their clustering on large scales. But problems arise with this picture if the initial perturbations are traced back to the earliest moments of the Big Bang. For many kinds of dark matter, the presence of fluctuations at the epoch of cosmic recombination, when the cosmic microwave background radiation last interacted with matter, tends to generate anisotropy in the radiation which has not been observed; in schemes which thus far manage to avoid this problem, such as the cold dark matter model, the power spectrum emerging from the early universe is not consistent with the large scales of galaxy clustering. It has been difficult to arrive at a comprehensive scheme that satisfies all of the observational constraints^{1–3,38}.

Thinking would be very different if there were known to be a true instability in ordinary atomic matter at very low density which could spontaneously generate perturbations from scratch after cosmic recombination. If such an instability were to occur, it would take the central role in the early development of cosmic structure and avoid the problem of excessive microwave background anisotropy. Ideally, the instability would produce linear fluctuations with the right amplitude and spectrum to agree with those required empirically for the gravitational formation of large-scale structure.

Under appropriate conditions, thermodynamic disequilibrium in the spectrum of the background radiation field could create just such a dynamical instability in the matter. Here I describe how the instability works, demonstrate how it could create cosmic perturbations, consider how the radiative disequilibrium necessary to trigger it might be created and propose some observational probes.

Although the instability discussed here occurs entirely in ordinary baryonic matter, it requires an unusual feature in the spectrum of cosmic radiation which is not part of the standard model. A suitable source of fresh radiation is required, for example a massive unstable neutrino which decays into ultraviolet

or X-ray photons. The properties of this radiation determine the evolution of structure, and observationally the strongest constraints on the instability come from background radiation. The particle lifetime must be long enough that the decay radiation is not thermalized in the black-body microwave background, but creates a spectral feature in the ultraviolet background radiation that interacts with neutral H or He atomic resonances some time after recombination. The resulting instability in the neutral gas creates large-scale fluctuations in gas density, leading to the formation of structure in much the same way as primordial isocurvature fluctuations², but because the instability creates perturbations after recombination, it is no longer necessary to create fully formed protogalaxies in the very early universe. If the radiation is sufficiently intense, the instability can create perturbations of the type required to generate galaxies and large-scale structure. The particle properties for this to occur differ somewhat from other recent decaying-particle proposals^{4–7}, but lie within the range of physical plausibility. The radiation intensity required is surprisingly small by cosmic standards, and for some parameters may even be discouragingly difficult to detect with current techniques.

Doppler instability

Consider an atom bathed in isotropic radiation with spectral energy density $i(\nu)$. Assume that at frequency ν the spectrum has logarithmic slope $\gamma \equiv d \ln i(\nu) / d \ln \nu$ —that is, $i(\nu) \propto \nu^\gamma$. Now suppose the atom is moving through the background at velocity $\beta \equiv v/c$. To first order in β , the spectrum of the radiation in this moving frame, observed at an angle θ from the direction of motion, is⁸

$$i'(\nu, \beta, \theta) = i(\nu)[1 + \beta(3 - \gamma) \cos \theta] \quad (1)$$

For $\gamma > 3$, the monochromatic radiation brightness is actually smallest in the forward direction. This is because for $\gamma > 3$, the photon occupation number in phase space (a Lorentz invariant scalar) increases with frequency. In the moving frame, photons of a particular frequency coming from the forward direction have been boosted from lower energy where the occupation number is lower, so there are fewer of them than from the backward direction where it is higher. This ‘Compton-Getting effect’ is a familiar one; it has been proposed, for example, as an observational tool in searching for line features in the background radiation⁹.

A runaway increase in the atom’s velocity results if a spectral feature with $\gamma > 3$ occurs near an atomic resonance. (If $\gamma < 3$, the motion of the atom is damped by resonant scattering, a feature exploited in the laboratory technique of ‘Doppler cooling’^{10–12}.) I focus here on interactions of atoms with radiation through coherent resonance scattering^{13,14}. The classical cross-section for unpolarized scattering from an electronic oscillator with frequency ν_0 is given by

$$\sigma(\nu) = \sigma_T [\nu^4 / (\nu_0^2 - \nu^2)^2] \quad (2)$$

where $\sigma_T \equiv 8\pi r_e^2/3$ is the Thomson cross-section. Expanding this expression about the line centre ν_0 yields the classical damping profile, $\sigma = (2\pi/3)r_e^2/\delta^2$, where the classical electron radius $r_e \equiv e^2/m_e c^2 = 2.8 \times 10^{-13}$ cm and $\delta = (\nu - \nu_0)/\nu_0$ is the fractional displacement of the photon from line centre. In each resonance scattering event, a neutral atom coherently scatters a

* Permanent address: Astronomy and Physics Departments, FM-20, University of Washington, Seattle, Washington 98195, USA.

photon into a random direction, the directional probability distribution having a $1 + \cos^2 \theta$ dependence for unpolarized radiation. Because of the symmetry of the scattering, the scattered radiation carries, on average, no momentum, so that the atom absorbs a mean momentum $h\nu/c$ in the original photon direction (although in the centre-of-mass frame the photon energy is always exactly unchanged by the scattering).

For simplicity I consider here only this phase-coherent elastic interaction in the damping wings of the atomic line. Other quantum absorption and re-emission processes do not preserve phase coherence or photon number, and are therefore inelastic or dissipative, requiring a more comprehensive analysis. As these effects are only important in the narrow Doppler core of the line, they do not affect the bulk of the photons driving the instability in the regime of interest, so the essential features are well described by this simplified model.

Computing the force on a single moving hydrogen atom is then similar to the calculation of Thomson drag between free electrons and the cosmic background radiation¹⁵—but with a γ -dependent sign. The momentum transfer which a moving atom receives from scattering the background radiation is obtained by integrating the absorbed momentum flux from all directions. I assume for the moment that γ is constant over some narrow range of frequencies in the vicinity of the Lyman- α resonance at frequency ν_α . If we consider only photons farther than δ to one side of line centre we find that for $\delta < \gamma^{-1}$

$$\begin{aligned} \dot{\beta} &= - \int d\Omega \int d\nu \sigma(\nu) i'(\nu, \beta, \theta) \cos \theta / mc^2 \\ &= (8\pi^2/9) \beta r_e^2 \nu_\alpha i(\nu_\alpha) (\gamma - 3) mc^2 \delta \end{aligned} \quad (3)$$

yielding an exponentially growing velocity $\beta \propto e^{\omega t}$ with growth rate

$$\omega(\delta) = \omega_0 / \delta \quad (4a)$$

where

$$\omega_0 = (8\pi^2/9) r_e^2 \nu_\alpha i(\nu_\alpha) (\gamma - 3) / mc^2 \quad (4b)$$

and m denotes the atomic mass. Thus, free energy in the form of a $\gamma > 3$ spectral distortion near Lyman- α tends to get channelled at this rate, for photons at fractional offset of order δ , into amplifying motions of matter.

Collective effects

The instability of an extended medium (as opposed to the acceleration of a single atom) is affected by several collective effects which change the behaviour. Here the cosmic Doppler instability exhibits features not found in line-driven instabilities in stars^{16–19}, because of the way the radiation field is modified as the scattering matter moves around²⁰. The radiation field tends to become isotropic in the frame moving with the local matter velocity; as in a cavity with radiation trapped between moving mirrors, this is a first-order effect. In addition, the mean photon spectrum changes in response to multiple scatterings; the global expansion leads to a cosmic redshift; and departures from the expansion in the form of perturbations lead to local compressions and rarefactions in the radiation. The latter effects produce secular changes in the mean spectral shape which are especially pronounced near line centre. The combination of these effects ultimately limits the attainable peculiar velocity, and tends to give large-scale coherence to the induced velocities of matter³⁹.

Photons change their energy slightly when they scatter off atoms, leading to a steady degradation of the spectral feature. Consider a photon scattering off an atom with velocity β . If $\nu\Omega$ and $\nu'\Omega'$ are the pre- and post-scattering photon vectors, the energy change in a scattering is²¹

$$\frac{\nu'}{\nu} = \frac{1 - \beta\Omega}{1 - \beta\Omega' + (h\nu/\gamma mc^2)(1 - \Omega\Omega')} \quad (5)$$

where $\gamma = (1 - \beta^2)^{-1/2}$. For very small $\beta < h\nu/mc^2$ the photon losses are dominated by Compton recoil, an effect we can neglect here. For larger β , however, we have

$$\Delta\nu/\nu \approx \beta(\Omega' - \Omega) \quad (6)$$

which is the same as the energy shift produced by a moving mirror. Because of the symmetry of the scattering direction, to first order in β a scattered photon has on average the same energy as an incident one. Individual scattering events, however, lead to energy gain or loss, depending on the direction, and correlations between successive scattering events, as well as second-order effects, produce secular changes in the mean spectrum. For $\gamma > 3$, where radiation accelerates matter, most of the scatterings at a given ν in the centre-of-mass frame involve loss because of the preponderance of higher-energy photons coming from behind. Photon energy thus goes into matter motions until the spectrum is flattened to $\gamma < 3$. As scatterings are most frequent near line centre, this part of the spectrum is degraded quickly and tends to damp the instability. The damping occurs preferentially on small scales; for scales smaller than the diffusion length of the $\gamma < 3$ photons, the effect of the radiation is to make the matter move coherently, because it damps the motion of an atom relative to the immediately surrounding matter. In particular, the instability does not amplify thermal motions of atoms. (The usual approach to solving for the spectral change from scattering using the Boltzmann-Kompaneets equation^{22,23} cannot be used, as the matter velocities are spatially highly correlated.)

We can estimate the velocities resulting from the instability just from the available photon energy, as the acceleration continues roughly according to equation (3) until the photon spectrum is changed. (Although it is far from obvious, this growth rate approximately holds in spite of the collective effects.) For photons with given offset δ , the photon energy losses must balance the kinetic energy gains of the matter due to these same photons. Let $\langle \beta^2 \rangle$ denote the mean square velocity of the atoms; then the rate of change of kinetic energy density of the matter, $\frac{1}{2} n m \langle \beta^2 \rangle$ (where n denotes the number density of atoms), must equal the rate at which energy is extracted from the photons. Using equations (3) and (4) we find that the spectrum at offset δ changes at a rate

$$\frac{d(\delta \nu i(\nu))}{dt} \approx - n m \langle \beta^2 \rangle \omega(\delta) \quad (7)$$

Eventually the velocities grow to the point where the instability saturates because of the photon reaction.

If the universe were not expanding, the instability would continue until all the photons were exhausted. For a spectral feature of fractional bandwidth ϵ , the fraction extracted from each photon is of the order of ϵ so the final saturated velocity is

$$\langle \beta^2 \rangle \approx 4\pi \epsilon^2 \nu i(\nu) / n m c^3 \quad (8)$$

In an expanding universe, the photons are redshifted by the cosmic expansion at the same time as losing their energy to the motion of matter through the instability, so the efficiency is reduced. In the case where $\omega_0 > H$, where H is the expansion rate, the expansion can be ignored altogether, reducing to the static limit. In general, however, losses due to the expansion dominate a photon's energy loss so only a small portion of the energy (within a narrow spectral band near the line centre) is available for creating peculiar motions. The net efficiency is $\delta_{\text{crit}} \approx \omega_0/H$, the offset where the growth rate of the instability exceeds H . The mean square velocity $\langle \beta^2 \rangle$ grows until the feedback (7) is important for $|\delta| \approx \delta_{\text{crit}}$. Outside this band the expansion losses dominate, so each photon ends up losing a fraction $\sim \delta_{\text{crit}}$ of its energy to the instability, the rest of it redshifting away in the expansion. For the final matter velocity

we then have instead of equation (8),

$$\langle \beta^2 \rangle \approx 4\pi \delta_{\text{crit}} \epsilon \nu i(\nu) / nmc^3 \\ = 4\pi (8\pi^2(\gamma-3)/9) \epsilon [\nu i(\nu)]^2 r_e^2 / Hnm^2 c^5 \quad (9)$$

I will use this expression below to relate the spectral intensity of the radiation to the amplitude of induced matter velocities and hence to the creation of structure.

The scale on which the velocities appear is also influenced by collective effects. Photons with the smallest δ , although they contribute the fastest growth rate at first, saturate quickly to $\gamma < 3$ and resist small-scale instability. Because these saturated photons suppress the instability on small scales, and more photon bandwidth is available at larger offset, larger velocities occur on larger scales. Photons with $\gamma < 3$ damp small-scale perturbation growth, but on large scales they are effectively trapped and are advected along with the matter flow, so perturbation growth proceeds unimpeded, powered by photons with larger δ . On very large scales—larger than the distance photons of offset δ can diffuse in a growth time $\omega(\delta)$ —the photons that create the instability are also effectively trapped and (for $\gamma > 3$) act like a gas with negative compressibility. In this limit the medium thus displays a simple macroscopic instability for the growth of long-wavelength fluctuations.

Roughly speaking the net efficiency is

$$\delta_{\text{crit}} \approx (\nu i(\nu) / nmc^3)(r_e^2 nc / H) \quad (10)$$

and the final mean square peculiar velocity is

$$\langle \beta^2 \rangle_{\text{crit}} \approx \epsilon (\nu i(\nu) / nmc^3)^2 (r_e^2 nc / H) \quad (11)$$

Perturbations of order unity are generated on the corresponding length scale,

$$\lambda_{\text{crit}} \approx (c/H) \epsilon^{1/2} (\nu i(\nu) / nmc^3)(r_e^2 nc / H)^{1/2} \quad (12)$$

The factor $(r_e^2 nc / H)$ emerges as an opacity parameter for the instability. For typical situations shortly after recombination, this number is of order unity. The key dimensionless parameter is the radiation flux in units of the mass density, $(\nu i(\nu) / nmc^3)$; if this is large enough, then λ_{crit} can be of galactic scale and the instability produces larger-scale perturbations which eventually lead to large-scale structure.

Large-scale structure

The large-scale density perturbations above λ_{crit} grow gravitationally after the instability ends, so we can adapt previous work on primordial fluctuations in predicting the statistical properties of the galaxy distribution. The power spectrum of fluctuations produced by the saturated Doppler instability has not yet been computed with the same rigour as the primordial models, so this discussion is necessarily qualitative.

The instability amplifies spatial peculiar velocities on a wide range of scales until the total kinetic energy density of the motions reaches the value in equation (9). Suppose a typical region of size λ has centre-of-mass velocity $\langle \beta^2 \rangle_\lambda^{1/2}$. From continuity, this spectrum of velocities produces a power density spectrum that has typical r.m.s. density fluctuations, in regions of scale λ , of $\langle (\delta\rho/\rho)^2 \rangle_\lambda^{1/2} \propto \lambda^{-3/2-n/2} \propto \langle \beta^2 \rangle_\lambda^{1/2} / \lambda$. If $\langle \beta^2 \rangle_\lambda$ is independent of λ , then the power spectral index $n = -1$. The minimal plausible amount of large-scale power from the instability corresponds to independent random perturbations on the scale λ_{crit} , yielding white noise on larger scales ($n = 0$); because of the damping of small-scale velocities, the final slope probably lies in the range $-1 \leq n \leq 0$. (Because of global energy and momentum conservation, a break in the spectrum must in any case occur above the comoving horizon scale, $\lambda \approx 10^2 h^{-1} [(1+z_i)/1000]^{-1/2}$ Mpc, where z_i is the redshift, consistent with large-scale microwave background anisotropy constraints.)

A power-law spectrum of density fluctuations with spectral index $-1 \leq n \leq 0$ is desirable for matching observations of cos-

mic large-scale structure³⁸. For $n = -1$, the three-dimensional galaxy correlation function on large scales has the form $\xi \propto r^{-(3+n)} \propto r^{-2}$ as indicated by observations of the large-scale galaxy distribution^{24,25}; even $n = 0$ gives a better fit than the asymptotic $n = 1$ produced by cold dark matter. Additional problems with the usual cold dark matter spectrum at small scales (for example, not forming quasi-stellar objects early enough²⁶) are also more easily solved with a power-law spectrum. Because of the way the fluctuations are generated, appreciable small-scale concentration of the baryons relative to dark matter is also expected, as suggested by dynamical evidence. The qualitative features of the fluctuations produced here thus come close to what one would prefer on purely empirical grounds. In most primordial schemes, on the other hand, it is unnatural to generate a spectrum that after recombination has a suitable form for satisfying all of the observational constraints^{1,38}. Another sense in which the Doppler mechanism is a more natural source of such perturbations than the very early universe is that nonlinear lumps on the scale λ_{crit} , corresponding roughly to protogalaxy masses, arise spontaneously after recombination rather than during the inflationary era.

Streaming velocities

The mechanism of Doppler instability can operate with many forms of dark matter (including hot), many particle decay parameters and so on. It is useful, however, to estimate the minimum required radiation intensity in a simple model to determine whether the Doppler instability can dominate structure formation in a physically and observationally plausible model.

For a minimum estimate of the radiation intensity needed to generate the observed structure, we can normalize with velocity or density data. A clean test is to use linear perturbation theory to relate the initial gas velocities to the current galaxy velocities, avoiding the theoretically and observationally contentious details of what the coherence scale of the velocities actually is. Let us assume that the bulk of kinetic energy density from the instability occurs on the same comoving scales that dominate galaxy peculiar velocities today. To reduce the number of parameters, assume for definiteness a flat ($\Omega = 1$) universe with a fractional baryon density Ω_b , and with the rest of the density collisionless, pressureless and invisible. In linear theory, an initial velocity β_i in the baryon component introduced as a linear perturbation at redshift z_i leads asymptotically (for $z_i \gg \Omega_b^{-1}$) to gravitational perturbations with peculiar velocity in the same component today²⁷

$$\beta_0 = (2/5)(1+z_i)^{1/2} \Omega_b \beta_i \quad (13)$$

For typical parameters, the initial baryon velocity is comparable to the final baryon velocity.

Combining this with equation (9) and scaling appropriately from z_i to the present, $[\epsilon \nu i(\nu)]_i = (1+z_i)^4 [\epsilon \nu i(\nu)]_0$, $n_i = (1+z_i)^3 n_0$ and $H_i = (1+z_i)^{3/2} H_0$, where $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $8\pi Gnm/3 = \Omega_b H^2$, we find that the present-day energy density of the spectral feature that generated the fluctuations must be at least

$$\left[\frac{4\pi \epsilon \nu i(\nu)}{c} \right]_0 = 7 \times 10^{-4} \text{ eV cm}^{-3} \left(\frac{\epsilon_i h^5}{\gamma-3} \right)^{1/2} \\ \times \left(\frac{\Omega_b h^2}{0.025} \right)^{-1/2} \left(\frac{\beta_0}{2 \times 10^{-3}} \right) \left[\frac{(1+z_i)}{1,000} \right]^{-9/4} \quad (14)$$

to create present-day peculiar motions of $\sim \beta_0$. I have adopted for normalization the value of Ω_b preferred from standard nucleosynthesis. The assumed velocity, $\beta_0 c = 600 \text{ km s}^{-1}$, is typical of observational estimates of r.m.s. velocities from data on large-scale flows, galaxy redshift catalogues and the peculiar motion of our own Galaxy²⁸. For comparison in these units, the microwave background energy density is $aT_{\text{bb}}^4 = 0.25 \text{ eV cm}^{-3}$ and the critical density is $10^4 h^2 \text{ eV cm}^{-3}$. Although a present-day

instability is not consistent with observations, it is certainly possible to have a post-recombination instability consistent with known backgrounds (see below).

Spectral feature

Pure primordial black-body radiation has a monotonically decreasing photon occupation number $(e^{h\nu/kT} - 1)^{-1}$, which is why some new source of photons, such as a decaying particle, needs to be added to the standard Big Bang model to produce the energy in the spectral feature (line features arising during standard cosmic recombination²⁹⁻³¹ do not depart sufficiently from equilibrium to give an instability). Thus far, spectra have been computed for only a few choices of particle properties^{4,32}, including some (for example in Fig. 4 of ref. 4) that have both the intensity and the shape required to trigger an instability.

It is worth sketching one particular example of a way such features form. There are many situations where a decaying particle would create a spectral feature leading to instability, because $\gamma > 3$ features are generically created by changes in cosmic opacity. This is most easily demonstrated in a simple model where the only opacity is photoelectric absorption by atoms.

Suppose decaying particles emit monochromatic photons with frequency ν_{decay} , at a rate per comoving volume $l(t) \propto e^{-t/\tau_{\text{decay}}}$ where τ_{decay} is the particle lifetime. In the absence of absorption, photons that appear at time t_0 and at frequency ν_0 were emitted at an earlier time $t_{\text{em}}(\nu_0, t_0)$ given by the cosmic redshift factor, $a[t_{\text{em}}(\nu_0, t_0)]/a(t_0) = \nu_0/\nu_{\text{decay}}$. In the presence of photoabsorption, the spectrum can be derived analytically at frequencies for which re-emission can be neglected, so I will use this case to demonstrate the origin of a destabilizing spectral feature. Standard transfer theory gives $i(\nu_0, t_0) \propto l(t_{\text{em}})(\partial t_{\text{em}}/\partial \nu_0)P(\nu_0, t_0)$, where $P(\nu_0, t_0)$ denotes the probability that a decay photon emitted at time $t_{\text{em}}(\nu_0, t_0)$ is not absorbed by time t_0 : $P(\nu_0, t_0) = \exp[-\tau(\nu_0, t_0)]$, where the optical depth is given by

$$\tau(\nu_0, t_0) = \int_{t_{\text{em}}(\nu_0, t_0)}^{t_0} c dt n(t) \sigma(\nu[t, t]) \quad (15)$$

$\sigma(\nu[t, t])$ denotes the cross-section for photoelectric absorption, and n is the density of absorbing atoms. As we are mainly interested in the logarithmic slope of the spectrum $\gamma = \partial \ln i / \partial \ln \nu$, we write

$$\ln i(\nu_0) = \text{constant} - t_{\text{em}}/\tau_{\text{decay}} + \ln(\partial t_{\text{em}}/\partial \nu_0) - \tau(\nu_0, t_0) \quad (16)$$

In taking the logarithmic derivative, the first term vanishes, and the second has a small absolute value for $t_{\text{em}} \ll \tau_{\text{decay}}$ and can be neglected. The derivative of the third term depends on the expansion law; although it is normally positive ($=3/2$ for a matter-dominated expansion), it will not normally yield $\gamma > 3$ on its own except in models that include unusual equations of state or a cosmological constant. The last term, however, is generically positive and has a large absolute value for some frequencies, leading to an instability for $\partial \ln i(\nu_0)/\partial \ln \nu_0 \approx -\nu_0 \partial \tau(\nu_0, t_0)/\partial \nu_0 > 3$.

The effect can be illustrated in a particular situation where these approximations all hold fairly well: where the resonant frequency triggering the instability lies above the absorption edge for the opacity. Suppose for example that $h\nu_{\text{decay}}$ lies just above the 584 Å Lyman resonance of neutral helium at 21 eV. These photons suffer continuum absorption by neutral hydrogen, $\sigma = \sigma_0(\nu_{\text{edge}}/\nu)^3$ where $h\nu_{\text{edge}}$ is the hydrogen Lyman continuum edge at 1 Ryd = 13.6 eV. Assuming constant comoving density of neutral hydrogen and a matter-dominated expansion law, we obtain

$$\gamma \approx -\nu_0 \partial \tau(\nu_0, t_0)/\partial \nu_0 = c t_0 n_0 \sigma_0 \left(\frac{\nu_{\text{edge}}}{\nu_0} \right)^3 \left[3 - \frac{3}{2} \left(\frac{\nu_0}{\nu_{\text{decay}}} \right)^{3/2} \right] \quad (17)$$

If the universe is optically thick in absorption, then the slope is steep enough for an instability. The dominant effect occurs when the optical depth is of order unity, so that not too many photons are absorbed. The steep cutoff at low frequencies occurs simply because photons that originate at higher redshift are more easily absorbed; this generic effect occurs in a wide variety of contexts where cosmic opacity is decreasing.

Background radiation

Neutral atomic gas after recombination has almost no coupling to the microwave background; the low-frequency limit of the scattering, Rayleigh scattering, goes as $\sigma \propto (\nu/\nu_0)^4$ and after recombination $kT \ll h\nu_0/40$. This justifies the neglect of radiation drag in the above analysis. The minimal interaction of neutral matter with black-body photons allows fluctuations to be generated with unusually small microwave spectral distortions and anisotropy.

In addition to the direct radiation in the destabilizing feature considered previously, another interesting question is whether the generation of the spectral feature or the action of the instability itself would create observable distortions in the black-body spectrum. Limits on the distortions are placed at present³³ at $\sim 1\%$ of the peak intensity, which lies considerably above the minimal energy budget for the instability, so these models would be constrained by known data only if Compton losses of decay photons dominate other losses by a very large factor. In general we require both better data and better computations of the spectra for a meaningful test. If the moving gas becomes ionized, there is a 'pseudo-Compton' distortion produced by the peculiar motions; this effect is relatively small because the net efficiency of creating the motions (δ_{crit}) is much less than 1. In fact the optimal time to create fluctuations from scratch with the least energy input is shortly after they enter the horizon; the comoving energy density in initial motions is substantially less than the resulting comoving binding energy density of the created perturbations and final collapsed systems³⁴.

If early star formation creates enough dust to absorb the destabilizing radiation before it redshifts out of the ultraviolet, the energy appears in what is now the submillimetre portion of the spectrum³⁵. In this case its spectral signature, although possibly detectable by the FIRAS instrument on the Cosmic Background Explorer (COBE)³³, would be indistinguishable from many other possible sources, including the stars themselves.

In the absence of dust obscuration, the background spectral feature that caused the instability should appear today in the background radiation spectrum. For interactions with the neutral hydrogen Lyman resonance after recombination, the feature with $\gamma > 3$ should appear redshifted into the infrared, but at a wavelength less than $0.12(1+z_i) \mu\text{m} \leq 120 \mu\text{m}$. If the instability is responsible for galaxy formation and clustering, the minimum radiation energy density is given in terms of model parameters by equation (14), about $4\pi[\epsilon \nu i(\nu)]_0/c \geq 10^{-3} a T_{\text{bb}}^4$. Such a feature in cosmic background radiation could in principle be detected by the DIRBE experiment on COBE satellite^{36,37}, although such a faint background may be difficult to detect behind the galactic and zodiacal emission. $\square$

Received 26 October 1990; accepted 6 March 1991.

1. Peebles, P. J. E. & Silk, J. *Nature* **346**, 233-239 (1990).
2. Peebles, P. J. E. *Nature* **327**, 210-211 (1987).
3. Saunders, W. et al. *Nature* **349**, 32-38 (1991).
4. Fukugita, M. & Kawasaki, M. *Astrophys. J.* **353**, 384-398 (1990).
5. Sciamma, D. *Astrophys. J.* **364**, 549-554 (1990).
6. Sciamma, D. *Nature* **348**, 617-618 (1990).
7. Sciamma, D. *Mon. Not. R. astr. Soc.* **246**, 191-192 (1990).
8. Peebles, P. J. E. & Wilkinson, D. T. *Phys. Rev.* **174**, 2168-2168 (1968).
9. de Bernardis, P., Masi, S., Melchiorri, B. & Melchiorri, F. *Astrophys. J.* **357**, 8-13 (1990).
10. Ashkin, A. *Science* **210**, 1081-1088 (1980).
11. Nagourney, W., Yu, N. & Dehmelt, H. *Opt. Comm.* **79**, 176-180 (1990).
12. Cohen-Tannoudji, C. N. & Phillips, W. D. *Physics Today* 33-40 (October 1990).
13. Heitler, W. *The Quantum Theory of Radiation* (Oxford, 1954).

14. Jackson, J. D. *Classical Electrodynamics* (Wiley, New York, 1975).
15. Peebles, P. J. E. *Physical Cosmology* (Princeton University Press, 1971).
16. Owocki, S. P. & Rybicki, G. B. *Astrophys. J.* **284**, 337–350 (1984).
17. Lucy, L. B. *Astrophys. J.* **284**, 351–356 (1984).
18. Owocki, S. P., Castor, J. I. & Rybicki, G. B. *Astrophys. J.* **335**, 914–930 (1988).
19. Katz, J. I. *Astrophys. J.* **317**, 264–270 (1987).
20. Rees, M. J. *Mon. Not. R. astr. Soc.* **213**, 75P–81P (1985).
21. Sunyaev, R. A. & Zeldovich, Ya. B. *Ann. Rev. astr. Astrophys.* **18**, 537–560 (1980).
22. Zeldovich, Ya. B. & Novikov, I. D. *Relativistic Astrophysics: Volume 2, The Structure and Evolution of the Universe* (University of Chicago Press, 1983).
23. Rybicki, G. B. & Lightman, A. P. *Radiative Processes in Astrophysics* (Wiley, New York, 1979).
24. Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, 1980).
25. Maddox, S. J., Efsthathiou, G., Sutherland, W. J. & Loveday, J. *Mon. Not. R. astr. Soc.* **242**, 43P–47P (1990).
26. Efsthathiou, G. & Rees, M. J. *Mon. Not. R. astr. Soc.* **230**, 5P–11P (1988).
27. Hogan, C. J. & Kaiser, N. *Mon. Not. R. astr. Soc.* **237**, 31P–38P (1989).
28. Rubin, V. C. & Coyne, G. V. (eds) *Large Scale Motions in the Universe: A Vatican Study Week* (Princeton University Press, 1988).
29. Peebles, P. J. E. *Astrophys. J.* **153**, 1–11 (1968).
30. Krolík, J. *Astrophys. J.* **338**, 594–604 (1989).
31. Krolík, J. *Astrophys. J.* **353**, 21–23 (1990).
32. Lyubarsky, Y. E. & Sunyaev, R. A. *Astr. Astrophys.* **123**, 171–183 (1983).
33. Mather, J. et al. *Astrophys. J.* **354**, L37–L41 (1990).
34. Hogan, C. J. *Mon. Not. R. astr. Soc.* **202**, 1101–1126 (1983).
35. Bond, J. R., Carr, B. J. & Hogan, C. J. *Astrophys. J.* **367**, 420–454 (1991).
36. Hauser, M. G. et al. in *After the First Three Minutes*, Goddard Space Flight Center (in the press).
37. Wright, E. L. in *Proc. Texas-ESO-CERN Symposium*, New York Academy of Sciences (in the press).
38. Kashlinsky, A. & Jones, B. J. T. *Nature* **349**, 753–760 (1991).
39. Hogan, C. J. *Astrophys. J.* (submitted).

ACKNOWLEDGEMENTS. I am grateful for useful comments from M. Fukugita, J. P. Ostriker, M. J. Rees and H. Sato. This work was supported by grants from NASA and from the Alfred P. Sloan Foundation.

Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5*

Osamu Chisaka & Mario R. Capecchi

Howard Hughes Medical Institute, Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA

Gene targeting in mouse embryo-derived stem cells has been used to disrupt the homeobox gene *hox-1.5*. Mice heterozygous at the *hox-1.5* locus appear normal, whereas *hox-1.5*[−]/*hox-1.5*[−] mice die at or shortly after birth. These homozygotes are athymic, aparathyroid, have reduced thyroid and submaxillary tissue and exhibit a wide range of throat abnormalities. In addition, they often feature defects of the heart and arteries as well as craniofacial abnormalities. These deficiencies are remarkably similar to the pathology of the human congenital disorder DiGeorge's syndrome.

AN intriguing hypothesis is that a common set of genes, known as the homeobox genes, may be used in specifying the body plans of invertebrates and vertebrates alike. In *Drosophila* these genes, which include members of the Ultrabithorax and Antennapedia complexes, specify the identity of cells within each parasegment^{1–3}. The function of the corresponding genes in man and mouse is not known, but as the order of these genes on the chromosomes of *Drosophila*, man and mouse is the same^{4,5}, and because this gene order reflects the order of anterior boundaries of gene expression along the anteroposterior body axis of the early embryos of all three species^{6–8}, it seems likely that these genes have an equally important role during mammalian development.

In man and mouse this set of genes, known collectively as the Hox genes, contains more than 30 members^{9–11}. They are distributed in the genome in four separate linkage groups, Hox-1–4, on four different chromosomes. The four linkage groups are believed to have arisen during chordate evolution as the result of two duplications of chromosomal segments¹². Expansion of this gene complex may have had a critical role in the progression from invertebrates to vertebrates by supplying the necessary complexity to this network of genes to accommodate the development of our complex body plan.

The homeobox genes of *Drosophila* encode transcription factors that share a DNA binding motif^{13–15}. Molecular and genetic analyses have indicated that these genes act as master switches

directing the course of morphogenic development of each segment^{16–19}. As the human and mouse genes share similar homeobox sequences^{20–22}, the Hox proteins are thought to function also as transcription factors participating in the specification of regional information in the early mammalian embryo.

To determine the genetic function for some of the Hox genes in the mouse, we used gene targeting^{23,24} in pluripotent mouse embryo-derived stem (ES) cells to create mice with specific mutations in these genes. We describe here the phenotype of mice that are homozygous for the disrupted *hox-1.5* gene.

Hox-1.5 is one of the earliest of the Hox genes to be expressed during mouse embryogenesis^{25–29}. By *in situ* hybridization *hox-1.5* messenger RNA has been detected in presomitic mesoderm and in ectoderm of 7.5–8.0-day embryos. As embryogenesis progresses, the boundaries of *hox-1.5* expression sharpen and extend rostrally into the developing hind brain, anterior to the otic vesicles. At this stage, *hox-1.5* expression is strongest in neural ectoderm but can also be detected in prevertebral somites. By 12.5 days of gestation, after which the genesis of most organs in the mouse is well under way, *hox-1.5* expression is observed in a broad spectrum of organs and tissues including pharynx, aortic trunk, thyroid, spinal cord, spinal ganglia, lungs, stomach, spleen and kidneys.

Disruption of *hox-1.5* in ES cells

Figure 1a shows the targeting vector, pHox 1.5-N/TK1, TK2, used to disrupt the *hox-1.5* gene in ES cells. It contains 11.5 kilobases (kb) of mouse genomic sequence encompassing the *hox-1.5* locus, with the neomycin resistance (*neo*^r) gene from pMCIneoA (ref. 30) inserted into the *Bgl*II site within the *hox-1.5* homeobox domain²⁵. Flanking the mouse sequences are the thymidine kinase genes, TK1 and TK2, isolated from herpes simplex virus (HSV) type I and II respectively. Two different HSV TK genes were used to construct this vector in order to reduce the potential for intraplasmid recombination. The linearized targeting vector was introduced into ES cells by electroporation³⁰ and the cells then grown in medium containing G418 plus gancyclovir to enrich for transformants containing the *neo*^r gene targeted into one of the endogenous *hox-1.5* loci³¹. Southern transfer analysis of DNA from G418 and gancyclovir-resistant colonies identified ES colonies that carry targeted disruptions of *hox-1.5*. One of 41 G418, gancyclovir-resistant colonies tested was positive. Figure 1b shows Southern blots of

DNA from the parental ES cell line, CC1.2 (ref. 32), and from the targeted ES cell line, 51-8, that were hybridized with DNA that flanks the *hox-1.5* gene (probe A) and is not present in the targeting vector, as well as with a *neo^r* probe (probe B). Cell line 51-8 is heterozygous at the *hox-1.5* locus, containing one copy of the mutated, *neo^r* disrupted allele, and one copy of the normal allele (Fig. 1b). In addition, 51-8 DNA was analysed

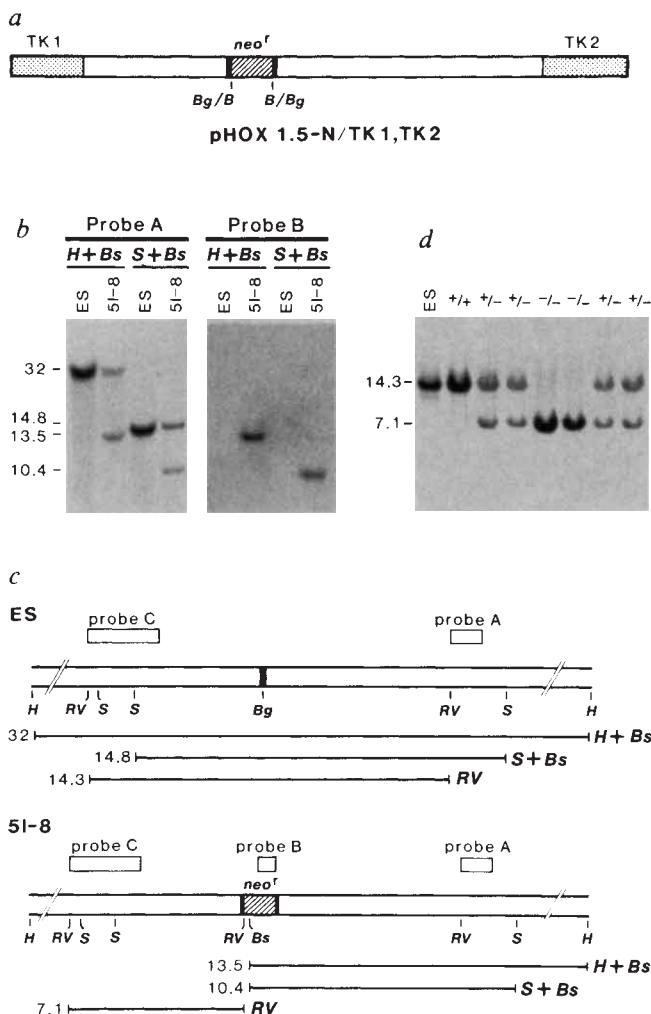


FIG. 1 Southern transfer analysis of ES cell line 51-8 containing a targeted disruption in the *hox-1.5* gene (a, b and c) and genotypic analysis of progeny derived from a *hox-1.5⁻/hox-1.5⁺* intercross (d). a, Structure of the targeting vector, pHOX 1.5-N/TK1, TK2, used to disrupt the *hox-1.5* gene in ES cells. b, Southern blot analysis of DNA from the parental ES cell line CC1.2 (ref. 32) and the targeted ES cell line, 51-8. The sizes of the DNA fragments are indicated in kb. The DNA samples were digested with *Hpa*I (H) and *Bst*BI (Bs), or with *Sac*I (S) and *Bst*BI, electrophoresed through agarose, transferred to nitrocellulose and hybridized with radioactively labelled probe as previously described³⁰. Probe A is a 3' *hox-1.5* genomic flanking probe, not present in the targeting vector (that is, a 1.2-kb *Eco*RV-*Cl*aI fragment, see map in c. Probe B is a 0.6-kb, *Pst*I-*Bam*HI *neo^r* fragment and probe C is a 5' *hox-1.5* genomic flanking probe (a 2.8-kb *Eco*RV-*Eco*RI fragment). c, Restriction map of the DNA fragments present in the parental and targeted ES cell lines CC1.2 and 51-8 respectively. d, Genotype, as determined by Southern transfer analysis, of 13.5-day-old embryos from interbreeding of *hox-1.5⁻* heterozygous mice. DNA was extracted from the yolk sac surrounding each embryo and analysed. DNA was digested with *Eco*RV and hybridized with flanking probe C. The genotype at the *hox-1.5* locus is indicated above each lane. +/+, Wild type; +/-, heterozygous at *hox-1.5*; -/-, homozygous for the *hox-1.5⁻* allele. ES represents a control lane of DNA from the parental CC1.2. The light hybridizing band migrating a little faster than the 14.3-kb fragment in the ES lane represents an *Eco*RV *hox-1.5* polymorphism present in the STO cell line used as a feeder cell line to propagate the ES cells. Bg, *Bgl*II; B, *Bam*HI; H, *Hpa*I; Bs, *Bst*BI; RV, *Eco*RV; S, *Sac*I.

using a *hox-1.5* genomic 5' flanking probe (probe C, data not shown) which showed no additional rearrangements of the targeted locus either 5' or 3' of the *neo^r* insertion.

Chimaeric mice were generated by microinjecting 51-8 cells (agouti coat colour) into blastocysts derived from C57Bl/6 mice (nonagouti coat colour), followed by surgical transfer of the blastocysts into the uteri of pseudopregnant mice. Seven chimaeric mice were born, all of which were males with >80% agouti coats. Test breeding of the chimaeric males to C57Bl/6 females showed that three of these animals transmitted the ES-derived agouti coat colour to their offspring. All of the offspring of one of the chimaeric males were agouti, whereas the others sired agouti offspring at a frequency of 20% to 30%. The frequency of generating germ-line chimaeras with the 51-8 cell line was comparable to that obtained with the parental cell line CC1.2.

Mice heterozygous for the *hox-1.5⁻* mutation are fertile and appear normal. The *hox-1.5* genotype of siblings, derived from intercrosses of *hox-1.5* heterozygotes, were evaluated by Southern transfer analysis of tail DNA obtained from newborn mice or from DNA extracted from the yolk sac surrounding each embryo. Figure 1d shows a representative analysis of such a litter at 13.5 days of gestation. All three genotypes expected were present.

Hox-1.5⁻/hox-1.5⁻ mice do not die *in utero*. Of 150 mid-to-late gestation embryos analysed for genotype, 36 were homozygous for the wild-type allele, 77 were heterozygotes and 37 were homozygous for the mutant allele. The apparent 1:2:1 segregation of *hox-1.5* alleles among these embryos suggests that *hox-1.5⁻/hox-1.5⁻* embryos are not preferentially aborted or absorbed.

Phenotype of newborn *hox-1.5⁻/hox-1.5⁻* mice

Mice containing two defective copies of the *hox-1.5* gene die at or shortly after birth, the longest period of survival being 12 h. Those that survived birth were noticeably purple and apparently died from problems associated with circulation of the blood and/or pulmonary failure.

In about half (20/38) of the newborn *hox-1.5⁻/hox-1.5⁻* mice the abdomen was bloated. After puncturing the abdominal wall, air rather than fluid was released; puncturing under water showed that air was present in the peritoneal cavity, in the stomach, and in the intestines.

Organ defects

The first feature apparent in embryos and newborn *hox-1.5⁻* homozygotes was an abnormal body shape. Figure 2 compares body shapes of control heterozygous (a, b and c) and homozygous (d, e and f) *hox-1.5⁻* mice at three different stages. Relative to their heterozygous litter mates, the homozygous *hox-1.5⁻* mice are much more rounded, their necks are shortened and their heads are characteristically tilted into their chests.

Examining sagittal and parasagittal sections of the 13.5- and 15.5-day embryos indicated that most *hox-1.5⁻/hox-1.5⁻* mice were athymic (compare, for example, Fig. 2b and e). This was confirmed by examining sections from six newborn mutant mice. Figure 3 shows a sagittal section of the throat region from a *hox-1.5* heterozygous and a homozygous newborn mouse. The thymus is obviously absent in the section from the *hox-1.5⁻/hox-1.5⁻* mouse. No thymus tissue was evident in any of the sagittal or parasagittal sections. In addition to the absence of thymus tissue, most of these newborn mice also lacked parathyroid tissue and the amount of thyroid tissue was greatly reduced. (In two mutant mice small amounts of thymic tissue could be detected, corresponding to the caudal tips of the thymic lobes.)

In the sections shown in Fig. 3a and b, it is also evident that the neck region of the *hox-1.5⁻/hox-1.5⁻* mouse is noticeably shorter relative to the control *hox-1.5⁺* heterozygous mouse. For example, note the distance from the heart and superior vena cava to the thyroid gland in the *hox-1.5⁻/hox-1.5⁻* mouse (Fig.

3b). At the same magnification of the throat, the thyroid, heart and superior vena cava from the *hox-1.5⁻* heterozygous mouse are not in the same photographic field (Fig. 3a).

The thoracic region of eight *hox-1.5⁻/hox-1.5⁻* newborn mice was examined microscopically and/or histologically and compared with that from normal heterozygous and wild-type litter mates. A range of cardiovascular defects was observed. Two of the mice were missing the right carotid artery, and the size and shape of the compartments of the heart were abnormal in seven *hox-1.5⁻/hox-1.5⁻* mice: this included dilation and thickening of the wall of the right atrium and left ventricle, as well as the right ventricle being smaller and having a thinner wall. Classical septal and transposition defects, including persistent truncus arteriosus, were not seen.

Figure 4 shows coronal sections from the thoracic region of a *hox-1.5⁻/hox-1.5⁻* mouse (a, b, d and f) and a control *hox-1.5⁻* heterozygous litter mate (c and e). Hypertrophy of both atria is prominent in the mutant mouse (compare Fig. 4c and d). Figure 4a and b show persistent patent ductus arteriosus. Three mutant mice but no control litter mates had this condition. Further, the aorta in this *hox-1.5⁻/hox-1.5⁻* mouse, as well as in three others, has a thin wall and is poorly developed (Fig. 4c and d). Stenosis of the aortic valve was also apparent. Figure 4e and f shows that this mouse has a defective pulmonary semilunar valve, containing two rather than the normal three cusps. This feature was observed in three *hox-1.5⁻/hox-1.5⁻* mice.

The hypertrophy of the atria and the enlargement of all of the chief veins, including the jugular veins, the pulmonary veins and the superior and inferior vena cava (Fig. 4a, b, d and f) are strong indicators of cardiovascular dysfunction. In addition, the most likely cause for early death of the *hox-1.5⁻/hox-1.5⁻* mice is cardiovascular dysfunction, although the details of the anatomical and physiological cause(s) for the cardiovascular dysfunction remain to be determined.

Although it is not apparent from the sections shown in Fig. 4, this *hox-1.5⁻/hox-1.5⁻* mouse, as well as four others, had greatly reduced amounts of submaxillary tissue relative to normal litter mates.

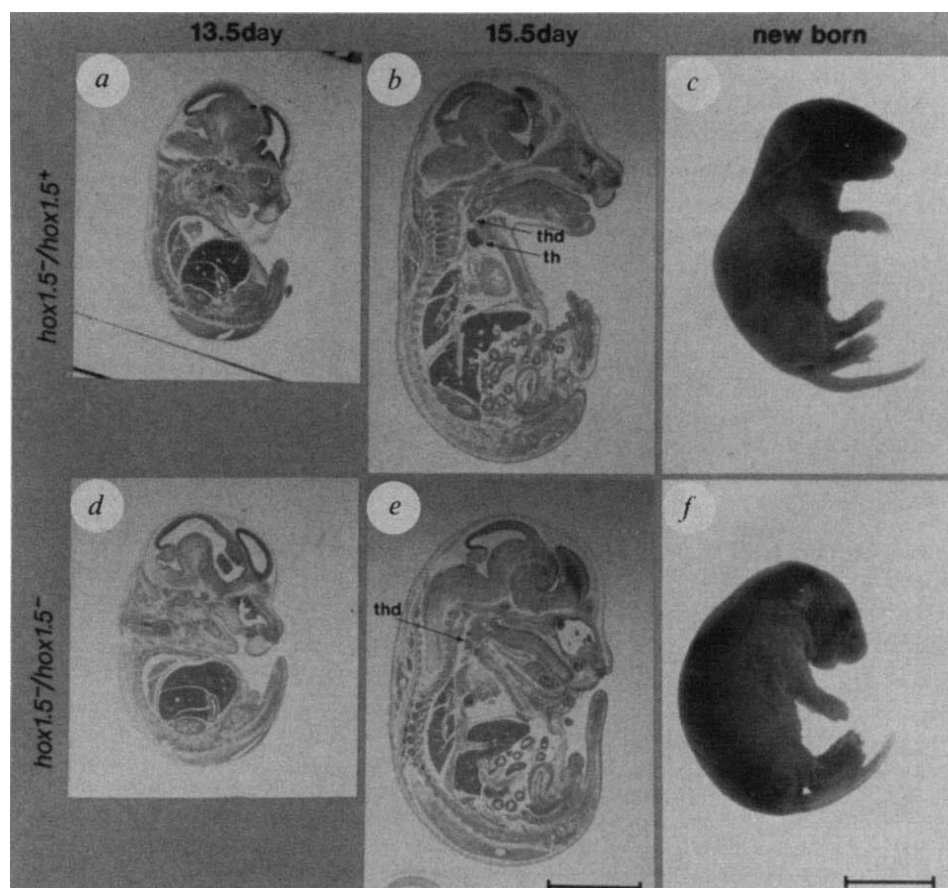
Skeletal defects

To examine further the shortened neck of *hox-1.5⁻/hox-1.5⁻* mice, trypsin/alkaline-cleared preparations of newborn mice were made and stained with alizarin red S and alcian blue to visualize the ossified bone and cartilage respectively³³. The number of cervical vertebrae was the same in heterozygous and homozygous litter mates, but the positioning of bones in the neck was different. In the mutant mouse shown in Fig. 5b, the entire thoracic region of the mouse is thrust towards the head, and the end of the clavicle is at the level of the axis rather than at the level of two cervical vertebrae caudal to the axis. Also, the larynx and trachea are positioned much closer to the cervical vertebrae than in control mice (see Fig. 5a-d). Detailed examination of the structure of the larynx revealed that the lesser horn (cornu minus) of the hyoid bone is absent in *hox-1.5⁻/hox-1.5⁻* mice (Fig. 5c and d). Also, the thyroid and cricoid cartilages are shorter and thicker in the mutant mouse. Facial features are also abnormal, due largely to the maxilla and mandible being shorter and altered in shape (see Fig. 5a and b). Measurements of the length of the mandible, relative to the height of the skull from four heterozygous and homozygous *hox-1.5* mice, indicated that this ratio was reduced by $6 \pm 2\%$ in the *hox-1.5⁻/hox-1.5⁻* mice relative to the control litter mates. A further example of facial malformations is that the zygomatic process of the squamosal bone is altered in the *hox-1.5⁻* homozygous mouse (Fig. 5e and f).

Bloated phenotype

About half of the *hox-1.5⁻/hox-1.5⁻* mice that survive birth have a bloated abdomen (Fig. 6b). To understand the cause of

FIG. 2 Comparison of body shapes of *hox-1.5⁻* heterozygous (a, b and c) and homozygous (d, e and f) mice; a, b, d and e show parasagittal and sagittal sections of 13.5- and 15.5-day-old embryos respectively; c and f are photographs of newborn mice. Before sectioning, embryos were fixed in Bouin's reagent, embedded in paraffin and regressive stained with haematoxylin and eosin (H and E) as described⁴⁴. The *hox-1.5⁻/hox-1.5⁻* mice and embryos are more rounded, their necks shortened and their heads tilted into their chest, relative to control, *hox-1.5⁻/hox-1.5⁺* heterozygous litter mates. thd, Thyroid; th, thymus. Scale bar in a, b, d and e, 5 mm; c and f, 6 mm.



the bloated phenotype, it is important to reexamine the throat region of the *hox-1.5⁻/hox-1.5⁻* mice. In Fig. 6*d-i* we compare sagittal sections of the throat from heterozygous and homozygous *hox-1.5⁻* newborn mice. As with the cleared preparations, structural differences in the larynx and trachea are apparent in these sections. The shapes of the thyroid and cricoid cartilages are altered in the *hox-1.5⁻* homozygous mouse, and the larynx is narrower. In addition to structural differences in the cartilages, the musculature in the throat is not as well organized as in the heterozygote, as is evident in the disordered array of muscle fibres in the lower part of the tongue (Fig. 6*f* and *i*), the connections of the muscle to the epiglottis (Fig. 5*e*, *f*, *h* and *i*) and the muscles surrounding the oesophagus (Fig. 6*d* and *g*). In the control animal, but not the *hox-1.5⁻* homozygote, the entrance to the oesophagus is constricted (Fig. 5, compare *d* and *e* with *g* and *h*), the epiglottis is pulled back from the entrance to the trachea (Fig. 5, compare *e* and *f* with *h*) and the soft palate is positioned to favour passage of air from the nasal passage to the trachea. Although the *hox-1.5⁻/hox-1.5⁻* mouse was breathing at the time of termination, the absence of control of the oesophagus and epiglottis may indicate a dysfunction in this musculature and/or neurons innervating these muscles. Similar structural malformations of the trachea and oesophagus are evident in the coronal sections of the *hox-1.5⁻/hox-1.5⁻* mouse shown in Fig. 4. The trachea and bronchi in the *hox-1.5⁻* homozygous mouse are much smaller than those of the control (Fig. 4*d* and *f*). Note in particular that the oesophagus in the *hox-1.5⁻/hox-1.5⁻* mouse (Fig. 4*d* and *f*) is not surrounded by a ring of muscle.

As previously mentioned, the stomach and intestines of *hox-1.5⁻/hox-1.5⁻* mice were found to be full of air. Because of difficulties with breathing, perhaps compounded by heart and arterial defects and possible dysfunction of the air passages in the throat, the *hox-1.5⁻/hox-1.5⁻* mice appear to pump air into their stomach and intestines. When the air pressure becomes sufficiently high, the lining of the stomach and/or intestine perforates, allowing leakage of air into the peritoneum. Similar pathology has been observed in human infants, where pumping of air into the stomach and intestines can result in perforation of their lining. This condition can be induced in human infants from the trauma associated with choking.

Structure of the pharyngeal arches

Many of the tissues and organs that are deficient in the *hox-1.5⁻/hox-1.5⁻* mouse are derived from the pharyngeal arches and/or pharyngeal pouches. For example, the thymus and parathyroid are derived from cells in pharyngeal pouch 3 and 4, the carotid artery is derived from pharyngeal arch 3 and so on. Thus, it may be significant that the pattern of pharyngeal arches in the 10.5-day embryos is altered. The relative position and perhaps the size of pharyngeal arch 4 is different in 10.5-day embryos of heterozygous compared with homozygous mice, and the demarcation between pharyngeal arch 3 and 4 is less well defined (Fig. 7*a-d*).

Sagittal, coronal and transverse sections of four 10.5-day embryos were examined. The rhombomeric pattern in the developing hind brain in *hox-1.5⁻/hox-1.5⁻* mutants appeared normal and all of the sensory and motor cephalic ganglia were also distinguishable and appeared histologically normal (data not shown).

Discussion

Gene targeting in mouse ES cells has been used to disrupt the *hox-1.5* homeobox gene. ES cells containing the disrupted *hox-1.5* gene were used to generate chimaeric mice that transmitted the mutant gene to their progeny, and intercrosses between mice heterozygous for the *hox-1.5⁻* mutation were used to generate *hox-1.5⁻/hox-1.5⁻* mice. These mice exhibit a complex phenotype with multiple defects in thymus, parathyroid, thyroid and submaxillary glands, facial cartilages such as the maxilla

and mandible, and the zygomatic process of the squamosal bone, hyoid, thyroid and cricoid cartilages, heart and great vessels issuing from the heart, carotid artery, soft palate, and connective and muscular tissues of the throat and tongue. Many of these defects are in tissues, cartilages and organs derived from all four pharyngeal arches and/or pouches. Thus, *hox-1.5* may play a critical part in the patterning of tissues in all of the pharyngeal arches and pouches. In addition, the spectrum of muscle deficiencies in the neck and tongue, as well as possible endothelial tissue deficiency in the aorta³⁴ may indicate that targets for *hox-1.5* include occipital myotomes, as well as the pharyngeal arch mesoderm that contributes to the muscle of the pharynx and larynx. Clearly, *hox-1.5* affects tissues and structures derived from multiple primary cell lineages.

Alternatively, or in addition, *hox-1.5* may be important for cephalic neural crest function. With the chick/quail chimaeric system³⁵⁻³⁹, it has been shown that the proper formation of the affected tissues and organs requires interaction of cephalic neural crest with cells in the pharyngeal arches and pouches. In addition to possible inductive processes involving interactions between neural crest and pharyngeal ectoderm, mesoderm and endoderm cells, neural crest also may contribute directly to the affected tissues by providing connective tissues derived from mesenchymal neural crest. Interestingly, microablation of premigratory cephalic neural crest in early chick embryos has effects very similar to those reported here for the *hox-1.5⁻/hox-1.5⁻* mouse, including loss of thymus, parathyroid and thyroid tissue, shortened necks, heart and arterial defects and alteration of pharyngeal arch architecture⁴⁰⁻⁴². The experiments by Kirby and her coworkers concentrated on removing cephalic neural crest caudal to the otic vesicles. The additional craniofacial abnormalities in the *hox-1.5⁻/hox-1.5⁻* mouse would argue that if the effect of the *hox-1.5⁻* mutation is at the level of cephalic neural crest, then the influence of *hox-1.5⁻* would also extend to neural crest cells rostral to the otic vesicles.

As *hox-1.5⁻* homozygous mice exhibit a specific phenotype it can be concluded that the other members of the *hox-1.5* subfamily (*hox-2.7*, *hox-3.6*, and *hox-4.1*) cannot fully complement the *hox-1.5* mutation. We can thus infer that the four clusters of closely related homeobox genes in vertebrates are not functionally redundant. Each member of the Hox complex may have a critical role in development. As a member of a network of genes, *hox-1.5* is likely to participate in cross regulation. One testable prediction of the above hypothesis is that the patterns of gene expression of other *hox* genes may differ in the *hox-1.5⁻/hox-1.5⁻* mouse. Although the phenotype of the *hox-1.5⁻/hox-1.5⁻* mouse is complex, it is nonetheless restricted with respect to the affected region of the mouse. Had the coupling between the Hox genes been extensive, then it might have been anticipated that the *hox-1.5⁻* mutation would have yielded an undecipherable early embryonic lethal phenotype. Instead, the *hox-1.5⁻/hox-1.5⁻* mice are born with a fairly well defined set of deficiencies.

The sequence of *hox-1.5* is most closely related to the *Drosophila* antennapedia complex genes *Zerknullt* (*zen*) and *proboscipedia* (*pb*). *Zen* is unique among this cluster of genes in that it seems not to be involved in the process of antero-posterior segmentation but rather in dorsal-ventral patterning⁴³. *Pb*, on the other hand, is involved in anteroposterior patterning and is required for differentiation of adult labial and maxillary structures⁴⁴. Loss of *pb* function transforms adult mouth parts to legs. The phenotypes associated with the loss of *hox-1.5* function in the mouse do not appear to involve transformations. Rather, as we have described, the loss of *hox-1.5* function most commonly affects the form of structures in a given region of the mouse and in this respect the *hox-1.5* gene seems to be required for regional refinement of pattern rather than for the establishment of pattern *per se*. The multiplicity of Hox genes in vertebrates may therefore be functionally correlated with developmental refinement of the more complex form

FIG. 3 *Hox-1.5*⁻/*hox-1.5*⁻ mice are usually athymic. Sagittal sections of the throat region from a *hox-1.5*⁻ heterozygous (a) and homozygous (b) mouse are shown. No thymus or parathyroid tissue could be detected in this *hox-1.5*⁻/*hox-1.5*⁻ mouse. After fixing and embedding in paraffin, the entire mouse was sectioned into 10- μ m sections and every fourth section was examined. The amount of thyroid tissue present in this mutant mouse is also greatly reduced. thd, Thyroid; th, thymus; svc, superior vena cava; h, heart. The width of a or b corresponds to 2.5 mm.

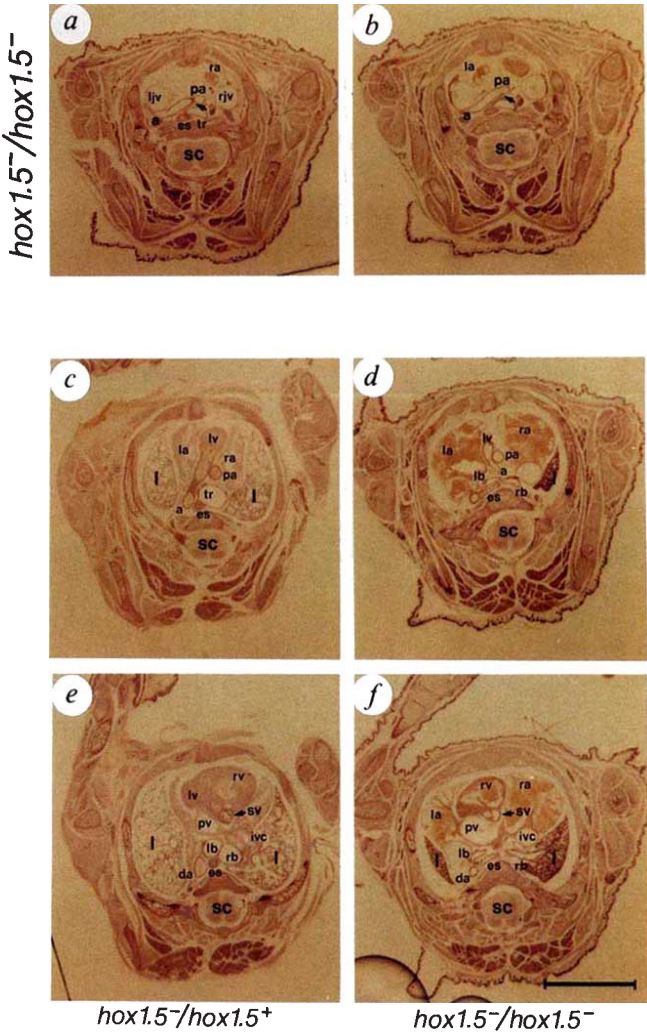
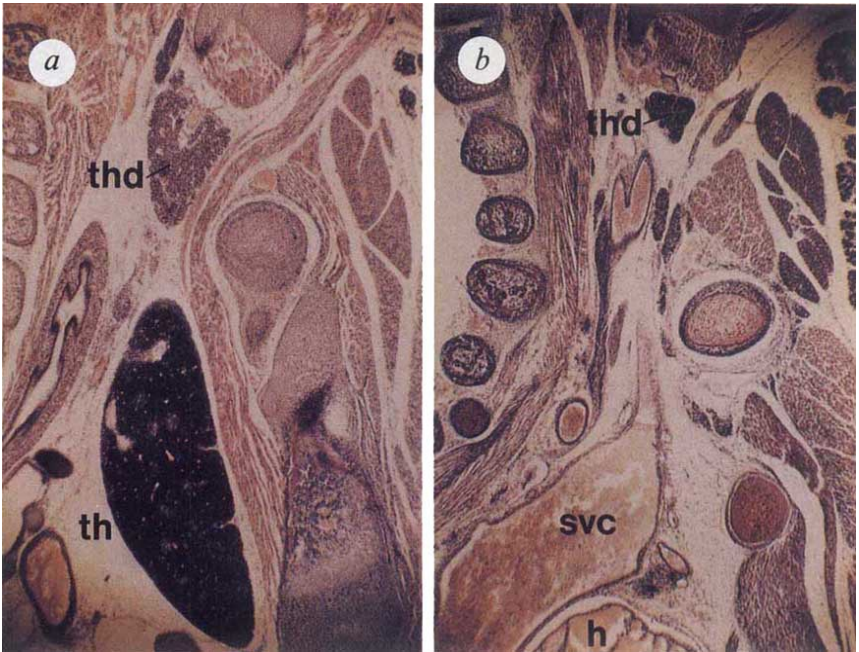


FIG. 4 Heart and arterial defects in the *hox-1.5*⁻/*hox-1.5*⁻ mouse. Coronal sections from the thoracic region of *hox-1.5*⁻ heterozygous (c and d) and homozygous (a, b, d and f) mice are shown; a and b show persistent patent ductus arteriosus (see arrows); c and d compare the aorta as it leaves the left ventricle in a control *hox-1.5*⁻ heterozygous mouse (c), and in a mutant *hox-1.5*⁻/*hox-1.5*⁻ mouse (d). Note that the aorta has a thin wall and is poorly developed in the mutant mouse; e and f show a coronal section through the pulmonary semilunar valve of a *hox-1.5*⁻ heterozygous (e) and homozygous (f) mouse. Note that this valve in the mutant mouse contains two, rather than the normal three, cusps, and the hypertrophy of the left and right atria as well as the enlargement of all the main veins in the mouse. a, Aorta; pa, pulmonary artery; da, descending aorta; la, left atrium; ra, right atrium; lv, left ventricle; rv, right ventricle; sv, semilunar valve; es, oesophagus; tr, trachea; rb, right bronchus; lb, left bronchus; sc, spinal cord; l, lungs; i, inferior vena cava; ijc, inferior vena cava; ijc, inferior vena cava; ijc, inferior vena cava. Scale bar, 3.5 mm.

of vertebrates.

The reported pattern of *hox-1.5* gene expression in mouse embryos does not correlate in any simple way with the pattern of defects we have observed in mice with mutated *hox-1.5* genes. The genetic function of *hox-1.5*, as defined here, is more spatially and temporally restricted than the gene expression pattern would predict. *hox-1.5* is expressed in the lungs, stomach, spleen and

kidneys²⁵⁻²⁹; however, no mutant phenotype can be detected in these organs. The lungs in *hox-1.5*⁻/*hox-1.5*⁻ mice may not function normally, but this seems likely to result from a secondary consequence of poor circulation of blood or premature death as judged by light microscopy. The size, shape and tissue quality of the lungs appear normal during development and in newborn mutant mice. What may be critical for correlating the

FIG. 5 Comparison of skeletons from *hox-1.5*⁻ heterozygous (*a*, *c* and *e*) and homozygous (*b*, *d* and *f*) mice. Newborn mice were cleared by treatment with trypsin and KOH and stained with alizarin red S and alcian blue to visualize the ossified bone and cartilage respectively³³. *a* and *b*, Head and neck regions. The whole thoracic region in the mutant mouse (*b*) is thrust towards the head, and the larynx is positioned closer to the cervical vertebrae compared with the control heterozygous *hox-1.5*⁻ mouse (*a*). In the lower jaw, the mandible also appears shorter and is thicker in the *hox-1.5*⁻/*hox-1.5*⁻ mouse. *c* and *d*, Detail of the hyoid bone and thyroid cartilage. In the *hox-1.5*⁻/*hox-1.5*⁻ mouse the lesser horn of the hyoid bone (cornu minus, *cm*) is missing. *e* and *f*, Zygomatic process of the squamosal bone (*zy*) is altered in shape in the *hox-1.5*⁻/*hox-1.5*⁻ mouse (*f*). *at*, Atlas; *ax*, axis; *clv*, clavicle; *cm*, cornu minus; *thdc*, thyroid cartilage; *zy*, zygomatic process of the squamosal bone. The bar represents 2 mm in *a* and *b*, but 1 mm in *c*–*f*.

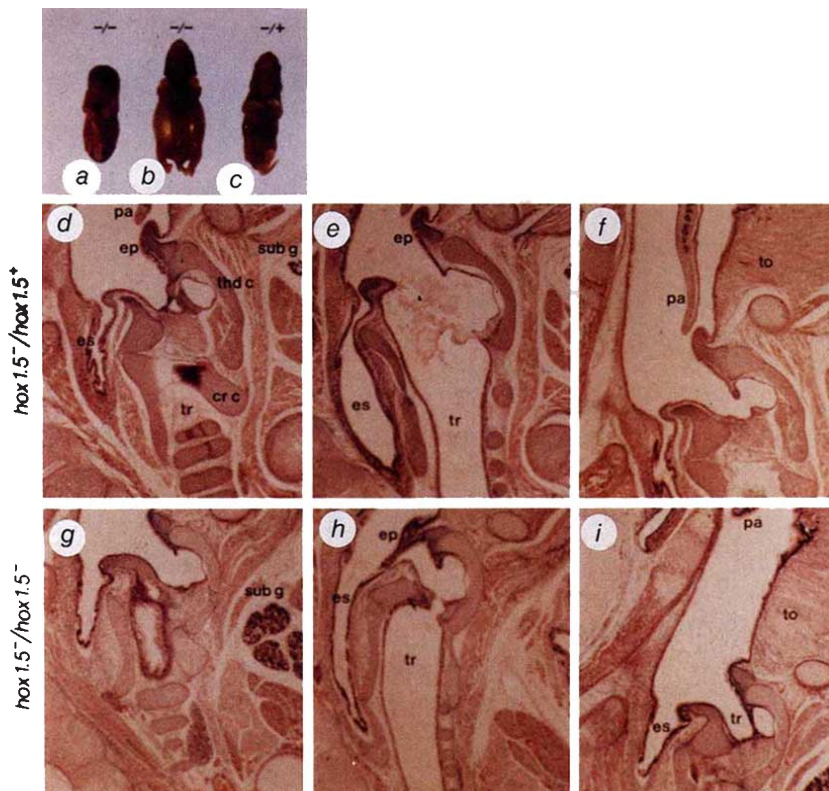
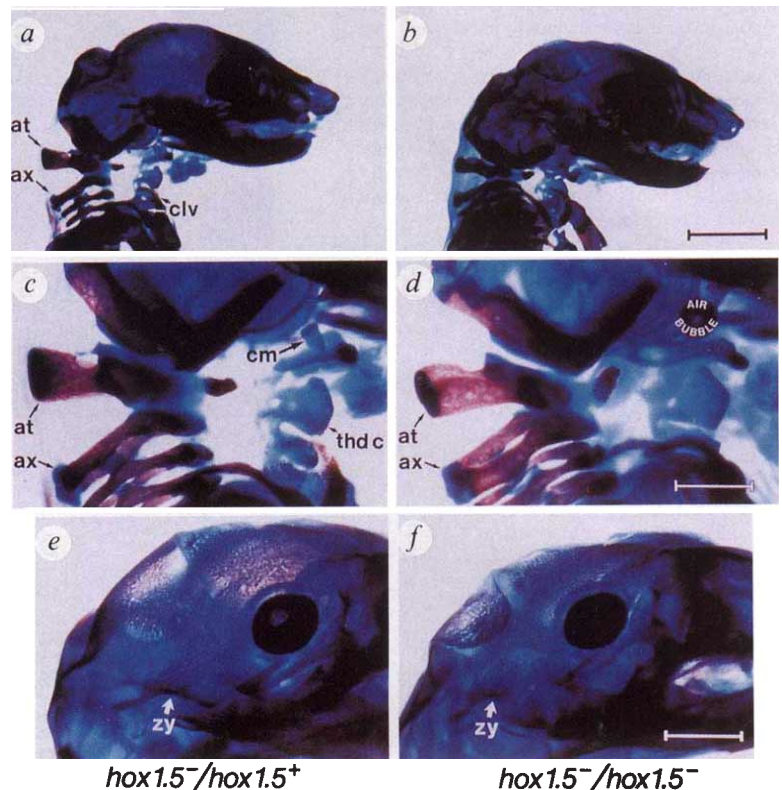


FIG. 6 Bloated phenotype and detail of the oesophagus and trachea. Photographs of newborn mice; *a*, a *hox-1.5*⁻/*hox-1.5*⁻ mouse that was dead at birth; *b*, a *hox-1.5*⁻/*hox-1.5*⁻ mouse that survived birth and developed a bloated abdomen; *c*, control *hox-1.5*⁻/*hox-1.5*⁻ heterozygous mouse. Sagittal sections that detail the throat region; *d*–*f* are sagittal sections from a heterozygous *hox-1.5*⁻ mouse; *g*–*i* are sagittal sections from a *hox-1.5*⁻/*hox-1.5*⁻ mouse. Both mice survived birth and were breathing at the time of termination. In the control *hox-1.5*⁻/*hox-1.5*⁻ mouse the entrance to the oesophagus is constricted (*d* and *e*), the epiglottis is pulled back (*d*, *e* and *f*) and the soft palate is positioned to allow air from the nasal passage to enter the trachea. In the *hox-1.5*⁻/*hox-1.5*⁻ mouse the entrance to the oesophagus is not constricted (*g*, *h* and *i*), the epiglottis is not pulled back (*h*) and the soft palate is truncated so that it does not favour the nasal or mouth passages. The organization of the muscle surrounding the oesophagus, connecting to the epiglottis, and in the back of the tongue are poorly organized in the *hox-1.5*⁻/*hox-1.5*⁻ mouse. *pa*, soft palate; *ep*, epiglottis; *subg*, submaxillary gland; *thdc*, thyroid cartilage; *cr*, cricoid cartilage; *es*, oesophagus; *tr*, trachea; *to*, tongue. The length of the bar in *d*–*i* represents 0.5 mm.

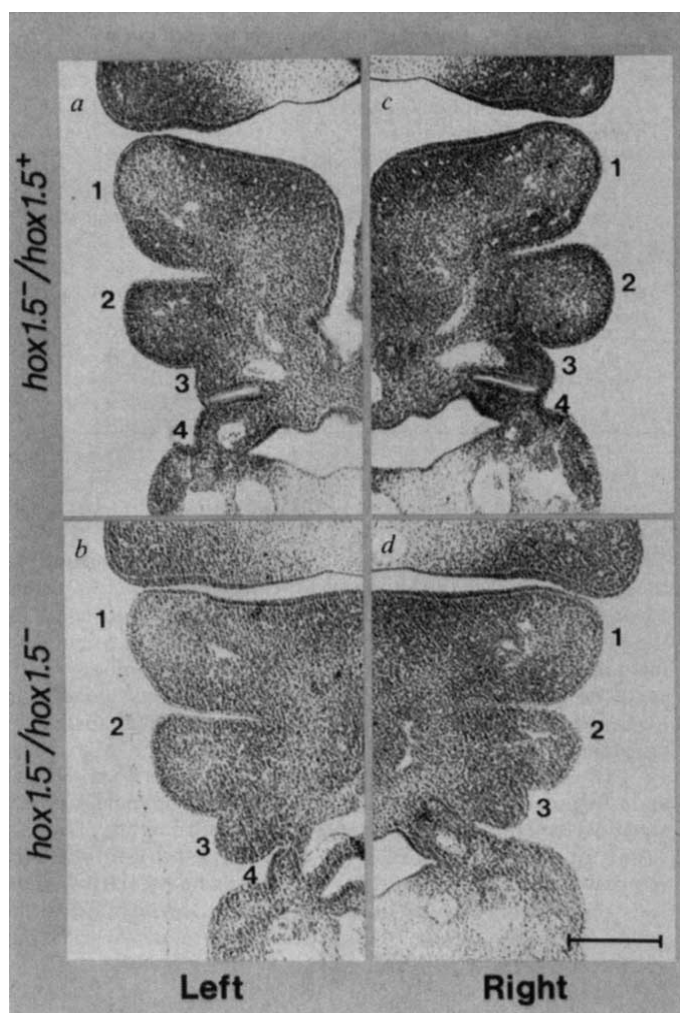


FIG. 7 Pharyngeal arches from *hox-1.5*⁺ heterozygous (a and c) and homozygous (b and d) mice. Coronal sections of 10.5-day-old mouse embryos illustrating the left (a and b) and right (c and d) pharyngeal arches. In the *hox-1.5*⁻/*hox-1.5*⁻ mouse, the left (b) and right (d) pharyngeal arch 4 is displaced. Also the demarcation between pharyngeal arches 3 and 4 is not as well defined in the *hox-1.5*⁻/*hox-1.5*⁻ mouse. Scale bar, 0.3 mm.

gene expression pattern with the mutant phenotype is the anterior portion of *hox-1.5* gene expression. A similar correlation was also observed in *int-1*⁻/*int-1*⁻ mice^{45,46}. Although *int-1* is expressed along the entire neural plate, from the mid brain to the caudal tip of the neural tube, phenotypic consequences of disrupting *int-1* were only apparent in the rostral portion of this expression pattern (that is, corresponding to the mid-brain and the anterior portion of the hind brain). The spinal cord, for example, was normal in the *int-1*⁻/*int-1*⁻ mice.

Humans with the congenital deficiency DiGeorge's syndrome are athymic, aparathyroid, have reduced thyroid tissue, bear craniofacial abnormalities and often are afflicted with cardiac and arterial defects⁴⁷. The coincidence of this set of deficiencies with those observed in the *hox-1.5*⁻/*hox-1.5*⁻ mouse is remarkable, although important differences exist between the human and mouse syndrome. DiGeorge's syndrome is autosomal dominant, whereas the mouse *hox-1.5*⁻ phenotype is autosomal recessive. DiGeorge's syndrome has been associated with deletions and translocations involving chromosome 22, whereas the human homologue to *hox-1.5* maps to chromosome 7. But as most patients with DiGeorge's syndrome are karyotypically normal⁴⁸, it is possible that the DiGeorge's syndrome phenotype could result from mutations in separate genes. Irrespective of whether or not the DiGeorge's syndrome phenotype of some patients can be attributed to mutations in the human *hox-1.5* gene, the similarity of phenotype in man and mouse argues for a common developmental pathway. In this context, the *hox-1.5*⁻ mouse may prove to be a useful model for DiGeorge's syndrome. □

Received January 31; accepted 7 March 1991.

- Lewis, E. B. *Nature* **276**, 565-570 (1978).
- Kaufman, T. C., Lewis, R. & Wakimoto, B. *Genetics* **94**, 115-133 (1980).
- Bender, W. *et al. Science* **221**, 23-29 (1983).
- Holland, P. W. H. & Hogan, B. L. M. *Genes Dev.* **2**, 773-782 (1988).
- Kessel, M. & Gruss, P. *Science* **249**, 374-379 (1990).
- Harding, K., Wedeen, C., McGinnis, W. & Levine, M. *Science* **229**, 1236-1242 (1985).
- Duboule, D. & Dolle, P. *EMBO J.* **8**, 1497-1505 (1989).
- Graham, A., Papalopulu, N. & Krumlauf, R. *Cell* **57**, 367-378 (1989).
- Acampora, D. *et al. Nucleic Acids Res.* **17**, 10385-10401 (1989).
- Duboule, D. *et al. Genomics* **7**, 458-459 (1990).
- Martin, G. *et al. Nature* **325**, 21 (1987).
- Kappen, C., Schughart, K. & Ruddle, F. H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5459-5463 (1989).
- McGinnis, W., Garber, R. L., Wing, J., Kurowia, A. & Gehring, W. J. *Cell* **37**, 403-408 (1984).
- McGinnis, W., Levine, M. L., Hafen, E., Kurowia, A. & Gehring, W. J. *Nature* **308**, 428-433 (1984).
- Scott, M. P. & Weiner, A. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3115-3119 (1984).
- Nusslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
- Akam, M. *Development* **101**, 1-22 (1987).
- Gehring, W. J. *Science* **236**, 1245-1252 (1987).
- Ingham, P. W. *Nature* **335**, 25-34 (1988).
- Carrasco, H. E., McGinnis, W., Gehring, W. J. & DeRobertis, E. M. *Cell* **37**, 409-414 (1984).
- McGinnis, W., Hirt, C. P., Gehring, W. J. & Ruddle, F. M. *Cell* **38**, 675-680 (1984).
- Scott, M. P., Tamkun, J. W. & Hartzell, G. W. *Biochim. Biophys. Acta* **989**, 25-48 (1989).
- Capecci, M. R. *Trends Genet.* **5**, 70-76 (1989).
- Capecci, M. R. *Science* **244**, 1288-1292 (1989).
- Gaunt, S. L., Miller, J. R., Powell, D. J. & Duboule, D. *Nature* **324**, 662-664 (1986).
- Fainsod, A., Awgulewitsch, A. & Ruddle, F. H. *Devl. Biol.* **124**, 125-133 (1987).
- Gaunt, S. L. *Development* **101**, 51-60 (1987).
- Gaunt, S. L., Sharpe, P. T. & Duboule, D. *Development Suppl.* **104**, 169-179 (1988).
- Gaunt, S. J. *Development* **103**, 135-144 (1989).
- Thomas, K. R. & Capecci, M. R. *Cell* **51**, 503-512 (1987).
- Mansour, S. L., Thomas, K. R. & Capecci, M. R. *Nature* **336**, 348-352 (1988).
- Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. *Nature* **309**, 255-256 (1984).
- McLeod, M. J. *Teratology* **22**, 299-301 (1980).
- Noden, D. M. *Ann. N.Y. Acad. Sci.* **588**, 236-249 (1990).
- LeLevré, C. S. & LeDouarin, N. M. *J. Embryol. exp. Morph.* **34**, 125-154 (1975).
- LeDouarin, N. *The Neural Crest* (Cambridge Univ. Press, 1980).
- Couly, G. & LeDouarin, N. M. *Development Suppl.* **103**, 101-113 (1988).
- Noden, D. M. *Devl. Biol.* **42**, 106-130 (1975).
- Noden, D. M. *Development Suppl.* **103**, 121-140 (1988).
- Kirby, M. L., Gall, T. F. & Stewart, D. E. *Science* **220**, 1059-1061 (1983).
- Bockman, D. E. & Kirby, M. *Science* **22**, 498-500 (1984).
- Bockman, D. E., Redmond, M. E. & Kirby, M. L. *Anat. Rec.* **225**, 209-217 (1989).
- Rushlow, C., Doyle, H., Hoey, T. & Levine, M. *Genes Dev.* **1**, 1268-1279 (1987).
- Pultz, M. A., Diederich, R. J., Gribbs, D. L. & Kaufman, T. C. *Genes Dev.* **2**, 901-920 (1988).
- Thomas, K. R. & Capecci, M. R. *Nature* **346**, 847-850 (1990).
- McMahon, A. P. & Bradley, A. *Cell* **62**, 1073-1085 (1990).
- McKusick, V. A. in *Mendelian Inheritance in Man* 916-917 (Johns Hopkins Univ. Press, 1990).
- Greenberg, F., Elder, F. B., Haffner, P., Northrup, M. & Ledbetter, D. M. *Am. J. hum. Genet.* **43**, 605-611 (1988).

ACKNOWLEDGEMENTS. We thank M. Allen, D. Harris, C. Lenz, E. Nakashima, R. Puryear and S. Tamowski for technical assistance, M. L. Kirby and E. M. Hammond for examining hearts from the *hox-1.5* mice, and D. Bramble for discussions.

Re-evaluation of the chemistry of dust grains in the coma of comet Halley

L. Mukhin*, G. Dolnikov*, E. Evlanov*, M. Fomenkova*, O. Prilutsky* & R. Sagdeev*†

* Space Research Institute, Profsoyuznaya strasse 84/32, Moscow 117810, USSR

† University of Maryland, College Park, USA

THE bulk chemical composition of the dust grains in the coma of comet Halley has been determined¹⁻¹¹ by examination of data from the PUMA 1 and 2 and PIA mass spectrometers on the Vega and Giotto missions. Although the bulk elemental composition of rock-forming elements (that is, excluding C, H, N and O) seemed to be close to solar (that is, similar to CI carbonaceous chondrites), the ion ratios of some of these elements, such as Mg^+/Si^+ and Fe^+/Si^+ , are rather different from those in CI chondrites. There has not been, however, a satisfactory investigation of the chemical composition of individual grains as a function of their mass. Here we re-evaluate the PUMA 1 and 2 data to perform such an analysis. We find that the compositions of heavy and light grains are very different, with light grains being magnesium-rich (silicon-deficient), whereas the mean Mg^+/Si^+ ratio in heavy grains is similar to CI chondritic. The marked difference in composition between light and heavy grains indicates that the origin of the two grain populations might be different.

The mass distribution of dust grains ranges over five orders of magnitude¹² (10^{-16} – 10^{-11} g). Obviously, the composition of the heaviest particles is likely to be most representative of the average chemical composition of the entire population of particles. It is not quite correct, however, to estimate the mass of the individual grains merely from the intensity of the ion signals in their mass spectra⁹. Instead, we should use independent data that characterize the impact of each dust particle with the target and the subsequent expansion of the plasma cloud. Such signals can be obtained from the target, accelerating grid, catcher electrodes and photomultiplier. (For a detailed description of the instrument, see ref. 13.) Cluster analysis was used to subdivide the spectra from each instrument into groups¹⁴ in accordance with these independent measurements. This approach made it possible to obtain the relative mass grouping of spectra; to define the absolute value of the dust particle mass for each group, the distribution of the number of spectra in the groups was compared with data on mass distribution obtained by the SP2 dust counter¹⁵. For PUMA 1 we also used the results of ref. 16: the lightest dust particle that gave a measurable spectrum had a mass of 5×10^{-17} g. For PUMA 2 we assumed the mass of the smallest registered particle to be the same as in the PUMA 1 experiment. Table 1 presents the calculated masses of dust particles for both instruments.

We used the independently derived particle masses as a weighting factor when calculating the average composition of the entire particle population. For our calculations, we summed the ion signals from the main elements in each spectrum and calculated the percentage of ions of each element. Averaging the percentage abundance of each element in the spectrum, rather than the absolute number of ions, reduces the uncertainty caused by variation of the ion signal in the spectra of particles of slightly different mass. We estimate that with this approach to determining the mean abundances the error is 25–30% for the more abundant elements. With less abundant elements such as Ca and Al, the error might reach 50%, because the ion amplitudes of these elements in the spectrum are small and often approach the value of the background. We checked for consistency by carrying out similar calculations for 1,383 spectra in the PUMA 1 instrument, first eliminating spectra with a high

TABLE 1 Mean dust particle mass for each group

Group number	PUMA 1		PUMA 2	
	Number of spectra	Mean mass (g)	Number of spectra	Mean mass (g)
1	155	8.28×10^{-17}		
2	312	2.13×10^{-16}	32	3.10×10^{-16}
3	87	5.35×10^{-16}	34	5.65×10^{-16}
4	329	1.30×10^{-15}	67	1.09×10^{-15}
5	249	3.34×10^{-15}	96	2.28×10^{-15}
6	144	9.33×10^{-15}	51	5.40×10^{-15}
7	181	2.99×10^{-14}	32	1.52×10^{-14}
8	200	1.29×10^{-13}	89	5.48×10^{-14}
9	293	5.18×10^{-13}	93	3.33×10^{-13}
10	60	2.24×10^{-12}	12	2.41×10^{-12}
11	21	5.0×10^{-12}	11	5.0×10^{-12}

background level. The results of the two calculations coincided within the indicated error.

Table 2 shows the calculated average ionic composition of the dust, with the allowance for the weighting factor for the PUMA 1 and 2 instruments. It also shows data calculated for the heavy and light fractions separately. It is evident that introduction of just the weighting factor, without any extra corrections, brings the calculated average ratios of the elements magnesium, silicon and iron in the comet dust into satisfactory agreement with their abundances in the sun: $Mg^+/Si^+ \sim Fe^+/Si^+ \sim 1$. The $S^+/Si^+ \sim 0.7$ ratio is also closer to the corresponding elemental ratio in CI than in previous papers. Conversion to elemental abundance using the yield factors of ref. 17 results in a Mg/Si ratio of ~ 0.3 , which is difficult to interpret. A possible explanation is that, at least for heavy particles, the ion yield factors derived in ref. 17 do not correspond to the actual processes in the PUMA instrument, and these factors seem to be mass-dependent.

Comparison of the Mg^+/Si^+ and $Fe^+/(Fe^+ + Mg^+)$ ratios in heavy and light grains shows that there is a significant difference in composition depending on particle mass (Table 2, Fig. 1). For example, the right-hand shoulders on Mg^+/Si^+ histograms of heavy and light particles (Fig. 1a and b) show that a considerable number of the light particles contain magnesium without accompanying silicon (possibly in the form of oxides,

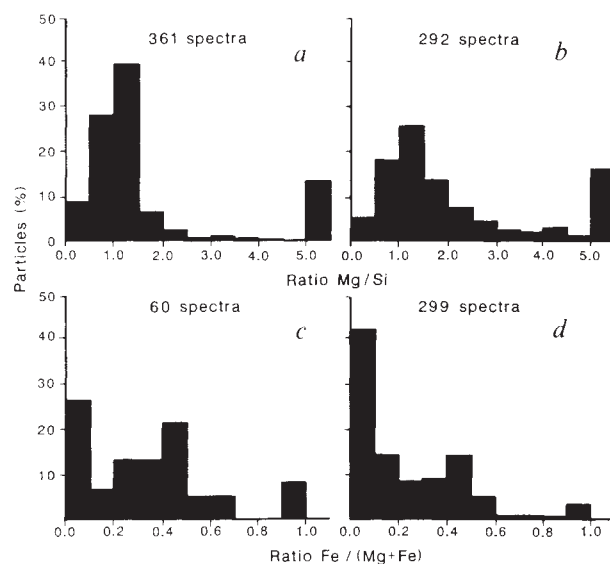


FIG. 1 Histograms showing the percentage of dust particles within various ranges of Mg^+/Si^+ or $Fe^+/(Fe^+ + Mg^+)$ ratio. All data are from the PUMA 1 instrument with a wide energy window. a, c, Heavy particles, mass $> 10^{-13}$ g. b, d, Light particles, mass $< 5 \times 10^{-16}$ g.

TABLE 2 Average ion abundances for PUMA 1 and 2 with weighting factor

	H	C	N	O	Mg	Al	Si	S	Ca	Fe
PUMA 1, wide*	2,937	548	74	491	100	19	94	65	14	88
PUMA 1, narrow	205	143	8.3	149	100	5.3	59	27	3.4	16
PUMA 2, wide	203	137	45	151	100	18	71	53	8.2	51
PUMA 2, narrow	117	50	13	48	100	13	57	49	7.1	29
PUMA 1, wide, heavy grains	3,316	579	77	530	100	19	97	70	14	97
PUMA 1, wide, light grains	258	252	18	103	100	17	47	27	9.5	22
C ₁ ²¹	2.5 × 10 ⁶	1,126	231	1,870	100	7.9	93	48	5.7	84

* 'Wide' and 'narrow' refer to energy windows.

hydroxides or carbonates). The mean Mg^+/Si^+ ratio for this group is much greater than 1. In heavy grains, this ratio is almost solar, and the number of grains containing magnesium without silicon is less than in the group of light grains. The heavy particles are more iron-abundant than the light particles (Fig. 1c and d).

The disagreement between our bulk results and those obtained previously can be readily interpreted by the fact that the number of light particles greatly exceeds that of heavy particles; introduction of the weighting factor therefore changes the bulk composition results.

The presence of a magnesium-rich phase in the light fraction dust is rather unexpected. Analysis of individual spectra from the PUMA 1 and 2 instruments that have a high Mg/Si ratio shows the probable presence in the dust of brucite, magnesium carbonates and, possibly, periclase. This assumption is supported to a certain extent by the discovery of the listed minerals in the carbonaceous chondrites^{18,19} and by thermodynamic evaluations²⁰, although more thorough analysis of the data is necessary.

The apparent difference in composition of the heavy and light particles can be explained in at least two ways: the difference may actually exist, or the effect may be due to the instrumental response to cometary particles of a specific mass. In our opinion, the first explanation is more probable. Comparison of the ionic abundance of rock-forming elements in CHON particles (those rich in C, H, O and N) and 'rock' particles (those rich in heavier elements) supports this argument. Figure 2 shows the distribution of Mg^+/Si^+ ratios in CHON and 'rock' particles: the histograms of Mg^+/Si^+ ratio in CHON particles are practically identical for light and heavy dust particles. In our opinion, this is proof the presence of a magnesium-rich phase in the light

dust fraction that does not occur in CHON particles, and this indicates that cometary dust particles of different origins are present. □

Received 30 July 1990; accepted 6 February 1991.

1. Kissel, J. et al. *Nature* **321**, 280–282 (1986).
2. Clark, B. C., Masson, L. W. & Kissel, J. *Proc. 20th ESLAB Symp. Eur. Space Ag. spec. Publ.* **250**, 353–358 (1986).
3. Jessberger, E. K. & Kissel, J. *Lunar planet. Sci.* **18**, 466–467 (1987).
4. Jessberger, E. K., Kissel, J., Fechtig, H. & Krueger, F. R. *Proc. Comet Nucleus Sample Return Workshop Eur. Space Ag. spec. Publ.* **249**, 27–30 (1986).
5. Mukhin, L. M. et al. *Lunar planet. Sci.* **18**, 674–675 (1987).
6. Clark, B. C., Masson, L. W. & Kissel, J. *Lunar planet. Sci.* **18**, 177–178 (1987).
7. Langevin, Y., Kissel, J., Bertaux, J. L. & Chassefiere, E. *Astr. Astrophys.* **187**, 761–766 (1987).
8. Jessberger, E. K., Kissel, J., Fechtig, H. & Krueger, F. R. in *Physical Processes in Comets, Stars, and Active Galaxies* (eds Hillebrandt, W., Meyer-Hofmeister, E. & Thomas, H. C.) 26–33 (Springer, Heidelberg, 1987).
9. Jessberger, E. K., Christoforidis, A. & Kissel, J. *Nature* **332**, 691–695 (1988).
10. Jessberger, E. K., Rahe, J. & Kissel, J. in *Origin and Evolution of Planetary and Satellite Atmosphere* (eds Atreya, S. K., Pollak, J. B. & Matthews, M. S.) 167–191 (University of Arizona Press, Tucson, 1989).
11. Mukhin, L. M. et al. *Lunar planet. Sci.* **20**, 733–736 (1989).
12. Jessberger, E. K. & Kissel, J. in *Comets in the Post-Halley Era* (ed. Newburn, R.) (Barnberg, Germany, 1989).
13. Balebanov, V. M., Scudrin, G. A. & Vorontsova, E. V. (eds) *Venus-Halley mission 88-93* (Louis-Jean, Paris, 1984).
14. Evlanov, E. N., Prilutsky, O. F. & Fomenkova, M. N. *XIX Meteorites Conf.* 67–68 (Moscow, 1990).
15. Mazets, E. P. et al. *Proc. 20th ESLAB Symp. Eur. Space Ag. spec. Publ.* **250**, 2, 3–10 (1986).
16. Sagdeev, R. Z. et al. *Space Res.* **25**, 840–848 (1987).
17. Kissel, J. & Krueger, F. R. *Appl. Phys.* **A42**, 69–85 (1987).
18. Zolensky, M. E., Barrett, R. A. & Pring, M. *Lunar planet. Sci.* **20**, 1253–1254 (1989).
19. Zinner, E. K., Cailliet, C. & Goresy, A. E. *Lunar planet. Sci.* **20**, 1245–1246 (1989).
20. Latimer, J. M., Schramm, D. N. & Grossman, L. *Astrophys. J.* **219**, 230–249 (1978).
21. Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197–214 (1989).

The latitude belts of solar activity as a consequence of a boundary-layer dynamo

G. Belvedere*, M. R. E. Proctor† & G. Lanzafame*

* Istituto di Astronomia, dell'Università, Viale Doria 6, Catania, Italy

† Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge, UK

HELIOSEISMOLOGY has provided much insight into the Sun's activity, allowing measurements of the thickness of the solar convection zone and the internal solar rotation rate as a function of latitude and depth. Solar activity is generally thought to result from dynamo action within the Sun, but there has been some debate as to whether this action occurs in the whole convection zone, at the base of the convection zone, or in the boundary layer between the convection and radiative zones^{1,2} (about $0.65 R_{\odot}$ to $0.7 R_{\odot}$). Here we point out that recent helioseismological data^{3–7} seem consistent with the last location. We present the results of calculations on a model in which dynamo action arises in a very thin ($0.05 R_{\odot}$) spherical boundary layer, which are consistent with the observations. In particular, this model provides an explanation of why there exists a latitudinal boundary on the solar surface, below which features such as sunspots migrate towards the equator whereas above it they migrate polewards.

The recent helioseismological data (K. G. Libbrecht, personal communication and refs 3–7) seem to indicate that the angular

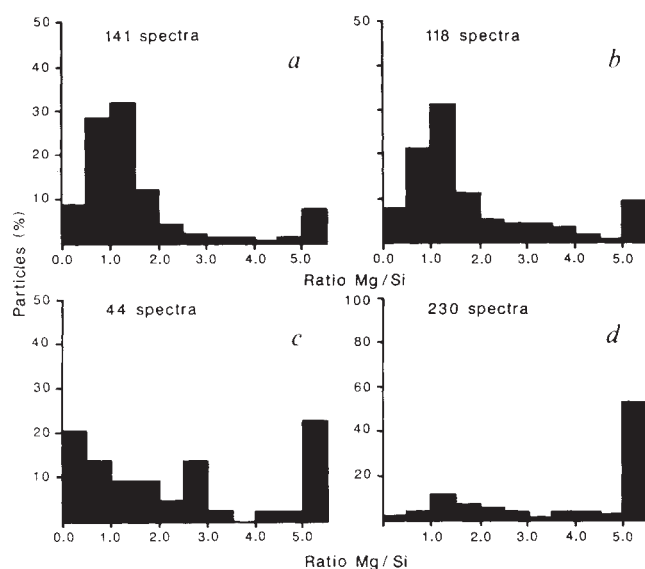


FIG. 2 Histograms showing the distribution of Mg^+/Si^+ ratio in four populations of dust particles having different compositions and masses: a, CHON particles, $m > 10^{-13}$ g; b, DHON particles, $m < 5 \times 10^{-16}$ g; c, 'Rock' particles, $m > 10^{-13}$ g; d, 'Rock' particles, $m < 5 \times 10^{-16}$ g.

velocity ω at the solar surface (a function of latitude) is maintained almost unchanged throughout the convection zone ($1 R_{\odot}$ to $0.7 R_{\odot}$), whereas at depths below this ($\approx 0.65 R_{\odot}$), rigid rotation occurs with $\omega_0 \approx 2.7 \times 10^{-6} \text{ rad s}^{-1}$, which is the surface value at latitude $\lambda_0 \approx 35^\circ$. Thus the latitudinally varying rotation rates above the convection zone converge to the intermediate value ω_0 below it. In our opinion (see also ref. 8), this seems to exclude the possibility of a dynamo, driven by radial shear, operating in the convection zone, but suggests that such a dynamo might operate in the boundary layer.

Considerations of convective overshooting motions shows that the dynamo parameter α , related to the helicity of turbulence, may be negative in the boundary layer⁹⁻¹¹. Interpolating the helioseismological data in the boundary layer itself, the radial angular velocity gradient $\partial\omega/\partial r$ should be positive for $\lambda < \lambda_0$ and negative for $\lambda > \lambda_0$. Thus, the well-known dynamo criterion^{12,13} based on the sign of the product $\Gamma = \alpha \cdot \partial\omega/\partial r$ allows propagation towards the equations of dynamo waves ($\Gamma < 0$) at lower latitudes ($\lambda < \lambda_0$) and poleward propagation ($\Gamma > 0$) at higher latitudes ($\lambda > \lambda_0$). This agrees with the observations shown by the latitude migration of different tracers of solar activity during the solar cycle. Indeed, sunspots and most faculae which belong to the equatorial activity belt ($\lambda < 35^\circ$) migrate towards the equator, whereas polar faculae, filaments and large-scale magnetic flux, which are characteristic of the polar activity belt ($\lambda > 40^\circ$ – 50°), migrate polewards.

Therefore, if the solar dynamo does indeed operate in the boundary layer, the two latitudinal belts of solar activity, which are distinguished by the opposite latitudinal migration of tracers, would appear as a natural and direct consequence of the internal rotation profile (the angular velocity gradient). Specifically, it is a consequence of the rigid rotation rate ω_0 below the convection zone being precisely the value at the surface at the critical latitude λ_0 that defines the boundary between the polar and equatorial belts. Although it is by no means the whole story, we feel that this may provide significant insight into the mechanism of the solar cycle. Our picture is supported by calculations on a model of such a boundary-layer dynamo. Dynamo models in cartesian geometry have been proposed before^{14,15} for the interface between the convection zone and the radiative interior, but those incorporate neither a realistic velocity field inside the Sun (compared with that inferred from acoustic oscillations), nor any discussion of the implications of a boundary-layer dynamo for surface activity.

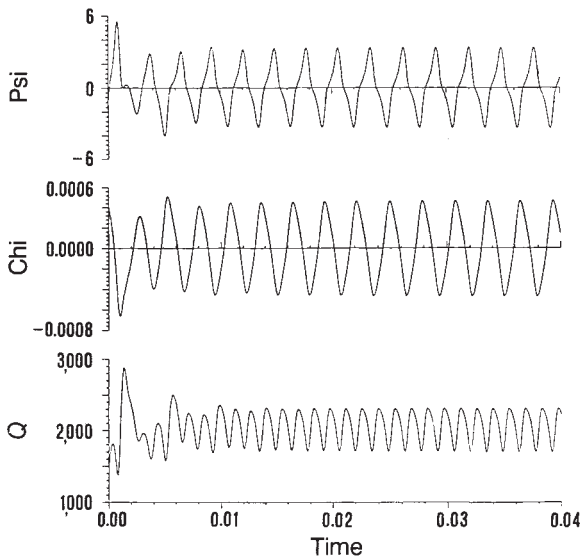


FIG. 1 Toroidal magnetic field (Psi), poloidal magnetic field (Chi) and zonal flow perturbation (Q) plotted as functions of time for a fixed intermediate latitude (45°) for $d=0.05 R_{\odot}$ and $D \sim -1$.

We have worked out a nonlinear α - ω dynamo model in a very thin ($0.05 R_{\odot}$) spherical shell. We have adopted an integrated representation of the spatial structure only in the radial direction (radial truncation), thus allowing full resolution in time and latitude of fields and flows. The latter allows us to construct butterfly diagrams, that show contours of toroidal magnetic field in a latitude-time plot, unlike previous studies¹⁶.

The mean field dynamo equations for the time evolution of the magnetic field $\mathbf{B} = B(r, \vartheta)\mathbf{\Phi} + \nabla \times (A(r, \vartheta)\mathbf{\Phi})$ where r, ϑ are spherical polar coordinates, $\mathbf{\Phi}$ is the unit vector in the azimuthal direction, $B\mathbf{\Phi}$ is the toroidal and $\mathbf{B}_p = \nabla \times A\mathbf{\Phi}$ the poloidal part of $\mathbf{B}$, are:

$$\frac{\partial A}{\partial t} = \alpha F(r, \vartheta)B + \eta_T \left[\nabla^2 - \frac{1}{r^2 \sin^2 \vartheta} \right] A \quad (2)$$

$$\frac{\partial B}{\partial t} = r \sin \vartheta \mathbf{B}_p \cdot \nabla \left[\frac{U(r, \vartheta)}{r \sin \vartheta} \mathbf{\Phi} \right] + \eta_T \left[\nabla^2 - \frac{1}{r^2 \sin^2 \vartheta} \right] B \quad (3)$$

Here αF is the usual α -effect, with F representing its spatial structure and α its magnitude. The quantity η_T is a turbulent diffusivity.

The dynamical influence of the magnetic field enters the model through its effect on the differential rotation $U(r, \vartheta) = u_0 + u$, where u_0 is a prescribed velocity field and u is a perturbation driven by the mean Lorentz force and subject to viscous damping. The simplest equation that encompasses these features is:

$$\rho \frac{\partial u}{\partial t} = \frac{1}{\mu_0} [(\nabla \times \mathbf{B}) \times \mathbf{B}]_{\Phi} + \rho \nu_T \left[\nabla^2 - \frac{1}{r^2 \sin^2 \vartheta} \right] u \quad (4)$$

where ν_T is a turbulent viscosity.

The model equations have been solved numerically with an explicit time-stepping model of DuFort-Frankel type with suitable boundary conditions. For full details see ref. 17, in which a similar model is given for the solar convection zone.

Assuming the differential rotation profile u_0 in the boundary layer as that given by interpolating the most recent helioseismological data, the results show periodic (dynamo wave-like) stable solutions (Fig. 1) with both equatorward and poleward migrating branches, as shown by the related butterfly diagram (Fig. 2). These solutions are found for dynamo number $D = \alpha \omega_0 d^3 / \eta_T^2 \approx -1$, where d is the thickness of the boundary layer and ω_0 is the surface angular velocity at latitude $\lambda \approx 35^\circ$.

The relatively small value of $|D|$ is a consequence of the very steep radial angular velocity gradient in the thin layer, because for the onset of dynamo action $|\alpha \cdot \partial\omega/\partial r|$ must exceed a characteristic value.

The relative latitude extent and magnetic field intensity of the

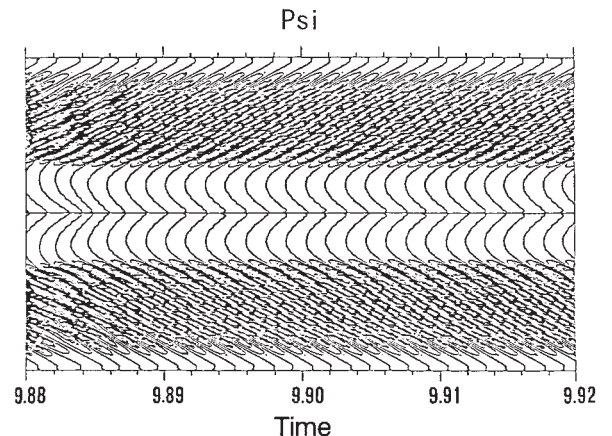


FIG. 2 Contours of the toroidal magnetic field (butterfly diagram) with equatorial and polar branches for the same run.

polar and equatorial branches is somewhat artificial, as it depends on the free choice of the latitudinal variation of the model parameter α , which has to satisfy only the requirement of antisymmetry with respect to the equator. Here we adopted $\alpha(\vartheta) \approx \sin \vartheta \cos \vartheta$, where ϑ is the polar angle, this giving more emphasis to the polar branch.

But a basic result of our numerical simulation is that horizontally propagating dynamo waves with equatorial and polar branches can exist even in a very thin layer. This was not obvious for a spherical shell, although not very surprising, as it has already been discussed in detail in a cartesian reference system^{14,15}. We conclude that the present numerical simulation supports the idea that the main surface activity features are intimately related to the angular velocity profile inside the

Sun through the working of the dynamo in the boundary layer.

There are several important facts that our model cannot explain, such as the well-known pattern of torsional oscillations¹⁸ and the 'extended' activity cycle¹⁹ (the overlapping pattern of magnetic field bands that migrates from high latitudes to the equator, beginning before the maximum of a given sunspot cycle and progressing during the next 18–22 yr). Furthermore, it remains to be understood why the internal rotation is rigid, why the radial gradient of angular velocity changes its sign at $\lambda \approx 35^\circ$ (see ref. 20 for a discussion of these issues), and why spots occur only for $\lambda < 35^\circ$. The answers to these questions require detailed studies of the interaction between rotation, convection and magnetic field in the solar interior. $\square$

Received 16 July 1990; accepted 18 February 1991.

1. Spiegel, E. A. & Weiss, N. O. *Nature* **287**, 616–617 (1980).
2. Spruit, H. C. & Van Ballegoijen, A. A. *Astr. Astrophys.* **106**, 58–66 (1982).
3. Harvey, J. W. in *Seismology of the Sun and Sun-like Stars ESA-SP 286* (ed. E. J. Rolfe) 55–66 (1988).
4. Brown, T. M. et al. *Astrophys. J.* **343**, 526–546 (1989).
5. Dziembowski, W. A., Goode, P. R. & Libbrecht, K. G. *Astrophys. J.* **337**, L53–L57 (1989).
6. Libbrecht, K. G. *Astrophys. J.* **336**, 1092–1097 (1989).
7. Morrow, C. A. in *Seismology of the Sun and Sun-like Stars ESA-SP 286* (ed. E. J. Rolfe) 91–98 (1988).
8. Belvedere, G. Proc. IAU Colloq. No. 121: 'Inside the Sun' (ed. G. Berthomieu & M. Cribier) 371–382 (Kluwer, Dordrecht, 1990).
9. Yoshimura, H. *Astrophys. J. Suppl.* **29**, 467–494 (1975).

10. Glatzmaier, G. A. *Geophys. Astrophys. Fluid Dyn.* **31**, 137–150 (1985).
11. Glatzmaier, G. A. *Astrophys. J.* **291**, 300–307 (1985).
12. Parker, E. N. *Astrophys. J.* **122**, 293–314 (1955).
13. Yoshimura, H. *Astrophys. J.* **201**, 740–748 (1975).
14. DeLuca, E. E. & Gilman, P. A. *Geophys. Astrophys. Fluid Dyn.* **37**, 85–127 (1986).
15. DeLuca, E. E. & Gilman, P. A. *Geophys. Astrophys. Fluid Dyn.* **43**, 119–148 (1988).
16. Weiss, N. O., Cattaneo, F. & Jones, C. A. *Geophys. Astrophys. Fluid Dyn.* **30**, 305–341 (1984).
17. Belvedere, G., Piddatella, R. M. & Proctor, M. R. E. *Geophys. Astrophys. Fluid Dyn.* **51**, 263–286 (1990).
18. Howard, R. & LaBonte, B. J. *Astrophys. J.* **239**, L33–L36 (1980).
19. Wilson, P. R., Alcock, R. C., Harvey, K. L., Martin, S. F. & Snodgrass, H. B. *Nature* **333**, 748–750 (1988).
20. Rosner, R. & Weiss, N. O. *Nature* **317**, 790–792 (1985).

Fast dynamo action in a steady chaotic flow

Andrew D. Gilbert

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Silver Street, Cambridge CB3 9EW, UK

THE observation of rapid variations in the Sun's magnetic field motivates the search for 'fast dynamos'^{1–3}—flows of highly conducting fluid that amplify magnetic fields on the typically rapid timescales of convection, rather than the longer timescales of diffusion. Certain helical flows^{4–6} have been proposed as possible fast dynamos, but numerical studies^{7,8} of such flows have shown no conclusive evidence for this. Here I examine the evolution of a magnetic field in one such flow, which possesses a web of chaotic streamlines mingled with tubes of regular streamlines⁹. In the case of no magnetic diffusion, I observe intense stretching and folding of the magnetic field in the chaotic regions of the flow. The folding brings together field that is largely aligned in the same direction, and the average field in a chaotic region therefore grows exponentially with time. This provides evidence for fast dynamo action, the main effect of weak diffusion being to average the field locally^{10–13}, and indicates that smooth, steady chaotic flows can be fast dynamos.

The existence of magnetic fields in the Sun and other astrophysical objects is broadly explained by dynamo theory, the amplification of magnetic field by flows of conducting plasma¹⁴. But the solar magnetic field is also very active; its dipole component reverses sign twice during each 22-year solar cycle, and features such as sunspots evolve within weeks. From the point of view of dynamo theory, it is interesting to compare this rapid evolution with other timescales in the Sun. The timescale of the motion of the convecting plasma is about a week, whereas the diffusive timescale² is roughly 10^6 years. The diffusive timescale is long because the plasma has a high electrical conductivity, and so the electric currents that support magnetic fields decay slowly. Clearly the Sun's field is evolving on the rapid convective timescale, and this motivates the search for models with this 'fast dynamo' behaviour. Fast dynamos may be contrasted with 'slow' dynamos, in which the field grows on a longer timescale

that is either diffusive or lies somewhere between the convective and diffusive timescales.

The kinematic dynamo problem is most tractable, when the magnetic field is taken to be so weak that the effect of the Lorentz force on the fluid flow may be neglected. The field then evolves according to the induction equation:

$$\partial_t \mathbf{B} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B} \quad (1)$$

in a given flow field $\mathbf{u}(\mathbf{x}, t)$. The magnetic diffusivity of the fluid is η and the electrical conductivity is $1/\eta$. A measure of the conductivity is the magnetic Reynolds number, defined by $R_m = UL/\eta$, where U is a typical velocity and L a typical length scale

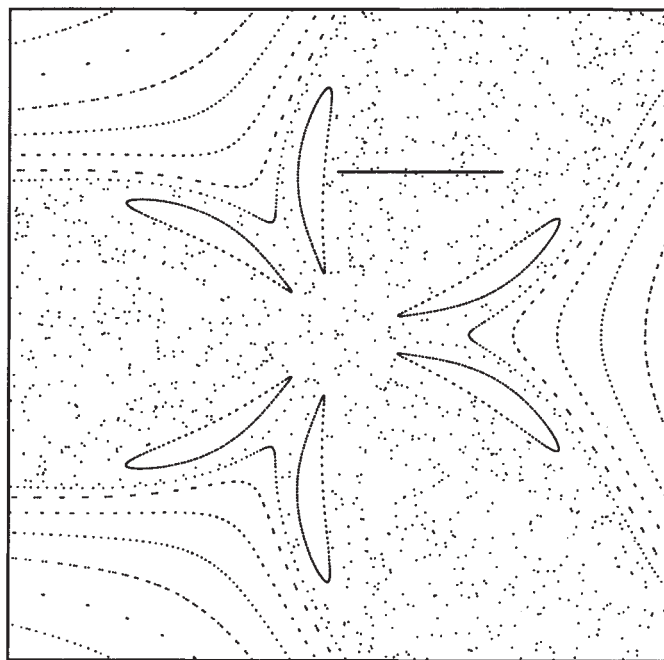
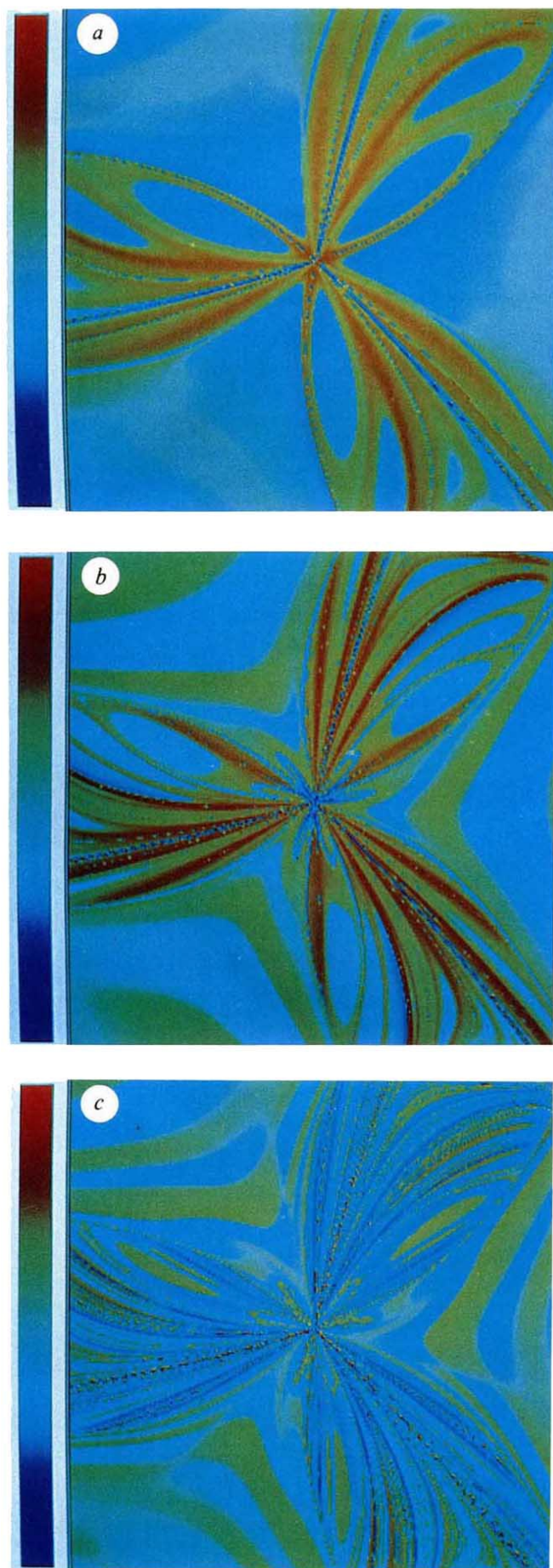


FIG. 1 Poincaré section of the ABC flow with $A=B=C=1$. The surface is centred at $(1, 1, 1)\pi/4$, where it intersects a separatrix, and extends from this point by $\pm(-1, 1, 0)0.15/\sqrt{2}$ in the horizontal direction on the page and by $\pm(-1, -1, 2)0.15/\sqrt{6}$ in the vertical direction. The horizontal line marked is used in calculating the average radial field, $\bar{B}_r$.



of the flow; R_m is also the ratio of the diffusive timescale, L^2/η , to the convective timescale, L/U . In the Sun $R_m \approx 10^8$, and it is very large in many other astrophysical objects². For a flow of any complexity, and a small value of η , dynamo action will generally take place¹⁵ and magnetic field will grow as $\exp(p(\eta)t)$. The flow field $\mathbf{u}$ is called a fast dynamo¹⁻³ if the growth rate remains on the convective timescale as $\eta \rightarrow 0$, so that $p(\eta)L/U$ remains positive and bounded away from zero. If, however, $p(\eta)L/U$ tends to zero, then the dynamo is slow.

Most of the dynamos that have been analysed in detail are slow, and there is so far no mathematical proof that fast dynamo action can occur for a smooth flow in bounded or periodic three-dimensional space. At present, therefore, fast dynamo research is aimed at proving that fast dynamos exist, rather than modelling the Sun in detail. One approach is to study fast dynamo models based on discrete mappings^{10,11,16,17}; these are simple enough to allow detailed analysis, but their relevance to realistic flows is unclear. There is strong numerical evidence for fast dynamo action in unsteady flows^{10,16,18}. Less is known about steady flows, although it has been suggested that these are typically fast dynamos¹⁷ and evidence for fast dynamo action has been found in certain idealized models¹⁹⁻²³. The steady flows that have attracted most attention belong to the *ABC* family⁴⁻⁶:

$$\mathbf{u}(x) = (B \cos y + C \sin z, A \sin x + C \cos z, A \cos x + B \sin y) \quad (2)$$

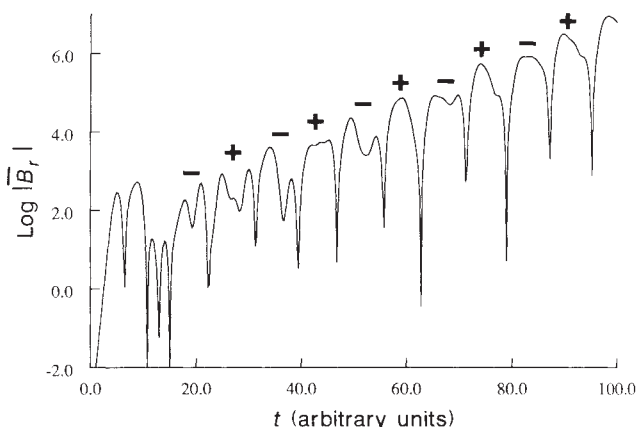
given by three scalar parameters, A , B and C . These periodic flows are exact three-dimensional solutions of the Euler equations for inviscid flow and generally have a complex streamline topology⁴⁻⁶. The flows have been used extensively in theoretical studies as they possess helicity, which is conducive to dynamo action¹⁴. For general values of the parameters an *ABC* flow has non-integrable streamlines that form a chaotic web⁹. An example is the case $A=B=C=1$, when the flow is highly symmetrical and possesses stagnation points connected by a network of separatrices²⁴. Figure 1 shows a Poincaré section of this flow: we plot points where streamlines cross a square surface embedded in space. At its centre the square intersects a separatrix, and the flow has threefold symmetry about this point. There are regions of chaotic streamlines, which imply intense stretching of the magnetic field, interspersed with islands in which the streamlines are quasi-periodic⁴⁻⁶.

Dynamo action in this flow has been studied numerically^{7,8} for values of R_m up to 550. The results show a window of dynamo action for $9 \leq R_m \leq 17.5$ and another window from $R_m \approx 27$ up to the limit of the computations. The numerical results perhaps suggest fast dynamo action, but larger computations are required to extrapolate to the limit of very large R_m with any confidence. The growing magnetic fields are concentrated in thin cigar-like structures extending along the separatrices connecting stagnation points.

Here I study the evolution of magnetic field in this *ABC* flow when there is no magnetic diffusion ($\eta = 0$), and field lines are 'frozen' in the fluid, being repeatedly stretched and folded by the flow. Now consider including very weak diffusion, $0 < \eta \ll 1$, which smooths field locally and dissipates magnetic energy. The effect of diffusion depends on the nature of folding in the flow; if the folding is constructive in that it brings together field pointing mostly in the same direction, diffusion dissipates relatively little magnetic energy and simply smooths out fine structure in the field. Thus the question of fast dynamo action for very weak diffusion becomes a question of whether there is

FIG. 2 Magnetic field evolving from the initial condition of equation (3) at times *a*, $t = 30$, *b*, $t = 45$ and *c*, $t = 70$. The field is shown at 256×256 points on the square surface of Fig. 1. The component of field, B_r , radial to the separatrix (and resolved perpendicular to streamlines), is represented, with red corresponding to positive values and blue to negative values.

FIG. 3 Average radial component of magnetic field, $\bar{B}_r$, averaged along the horizontal line marked in Fig. 1, plotted against time in log-linear coordinates. The sign of $\bar{B}_r$ is marked above peaks in the curve.



2. Zeldovich, Ya. B., Ruzmaikin, A. A. & Sokoloff, D. D. *Magnetic Fields in Astrophysics* (Gordon and Breach, New York, 1983).
3. Moffatt, H. K. *Nature* **341**, 285–286 (1989).
4. Arnold, V. I. *C. r. hebd. Séanc. Acad. Sci. Paris* **261**, 17–20 (1965).
5. Hénon, M. *C. r. hebd. Séanc. Acad. Sci. Paris* **262**, 312–314 (1966).
6. Dombre, T. et al. *J. Fluid Mech.* **167**, 353–391 (1986).
7. Arnold, V. I. & Korkina, E. I. *Vest. Mosk. Un. Ta. Ser. 1. Math. Mec.* **3**, 43–46 (1983).
8. Galloway, D. & Frisch, U. *Geophys. astrophys. Fluid Dyn.* **36**, 53–83 (1986).
9. Beloshapkin, V. V. et al. *Nature* **337**, 133–137 (1989).
10. Finn, J. M. & Ott, E. *Phys. Fluids* **31**, 2992–3011 (1988).
11. Bayly, B. J. & Childress, S. *Geophys. astrophys. Fluid Dyn.* **49**, 23–43 (1989).
12. Vishik, M. M. *Geophys. astrophys. Fluid Dyn.* **48**, 151–167 (1989).
13. Finn, J. M. & Ott, E. *Phys. Fluids* **B2**, 916–926 (1990).
14. Moffatt, H. K. *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge University Press, 1978).
15. Roberts, G. O. *Phil. Trans. R. Soc.* **271**, 411–454 (1972).
16. Bayly, B. J. & Childress, S. *Geophys. astrophys. Fluid Dyn.* **44**, 211–240 (1988).
17. Finn, J. M., Hanson, J. D., Kan, I. & Ott, E. *Phys. Rev. Lett.* **62**, 2965–2968 (1989).
18. Otani, N. *EOS Trans. Am. geophys. Un.* **69** (abstr.) 1366 (1989).
19. Arnold, V. I., Zeldovich, Ya. B., Ruzmaikin, A. A. & Sokoloff, D. D. *Zh. Eksp. Teor. Fiz.* **81**, 2052–2058 (1981); (Engl. transl.) *Sov. Phys. JETP* **54**, 1083–1086 (1981).
20. Soward, A. M. *J. Fluid Mech.* **180**, 267–295 (1987).
21. Gilbert, A. D. *Geophys. astrophys. Fluid Dyn.* **44**, 214–258 (1988).
22. Gilbert, A. D. & Childress, S. *Phys. Rev. Lett.* **65**, 2133–2136 (1990).
23. Finn, J. M., Hanson, J. D., Kan, I. & Ott, E. *Preprint UMLPR 90-015* (Univ. of Maryland, 1990).
24. Childress, S. & Soward, A. M. in *Chaos in Astrophysics* (ed. Buchler, J. R. et al.) 223–244 (Reidel, Dordrecht, 1985).

ACKNOWLEDGEMENTS. I am grateful to V. I. Arnold, S. Childress, U. Frisch, I. Klapper & H. K. Moffatt for valuable discussions. S. P. Cooper and the Atmospheric Dynamics Group of DAMTP generously made available the software package that was used to produce Fig. 2. This research was supported by Gonville and Caius College, Cambridge.

Calculating the rate of loss of information from chaotic time series by forecasting

David J. Wales

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, UK

DETERMINING whether time series of data from dynamical systems exhibit regular, stochastic or chaotic behaviour is a goal in a wide variety of problems. For sparse time series (those containing only of the order of 1,000 data points), the goal may simply be to discover whether the series are chaotic or not. Examples are case rates for infectious diseases¹ and proxy palaeoclimatic records from deep-sea cores². Sugihara and May³ have recently extended previous work⁴ aimed at distinguishing chaos from noise in sparse time series. Their approach is based on a comparison of future predictions of terms in the time series—derived using a data base of information from another part of the series—with the known terms. Here I present a method for estimating from such forecasting the largest Liapunov exponent of the dynamics, which provides a measure of how chaotic the system is—that is, how rapidly information is lost from the system.

The principal aims of the analysis of chaotic time series are to determine the Kolmogorov entropy (or K-entropy⁵) and fractal dimension that characterize the motion⁶. The Kolmogorov

constructive folding of magnetic field at zero diffusion; this connection can be made rigorous in certain idealized models^{10–13}. We shall look for constructive folding in an *ABC* flow; this approach has been used recently in a study of a steady, but non-smooth, chaotic flow constructed by patching together simpler flows²².

We use the highly symmetric initial field,

$$\mathbf{B}_0(\mathbf{x}) = (\sin z - \cos y, \sin x - \cos z, \sin y - \cos x) \quad (3)$$

and evolve it according to the Cauchy solution. It is also useful to resolve the field into components parallel, $\mathbf{B}_\parallel = (\mathbf{B} \cdot \mathbf{u})\mathbf{u}/u^2$, and perpendicular, $\mathbf{B}_\perp = \mathbf{B} - \mathbf{B}_\parallel$, to each streamline²². We calculate the magnetic field at various times on the square surface of Fig. 1, which is perpendicular to one of the separatrices in which the cigar-like structures are observed in ref. 8. The field has threefold symmetry about the separatrix that intersects the centre of the square, as can be seen in Fig. 2, which shows the field at times $t = 30, 45$ and 70 (arbitrary units). We show the component of field, B_r , radial to the separatrix (strictly this is the radial component of $\mathbf{B}_\perp$). Field pointing radially outwards is coded in red and field pointing radially inwards in blue. The strongest magnetic fields occur in the chaotic regions of the flow (see Fig. 1); here the field becomes increasingly fine-scaled and strong because of chaotic stretching and folding. In the integrable regions the field remains relatively large-scale and weak. The folding action tends to bring together like-signed field: in Fig. 2a and b the field is largely pointing radially outwards (red) whereas in Fig. 2c it is mostly inwards (blue). Thus the folding by the flow seems to be constructive and to produce a mean field (which oscillates in time) in the chaotic parts of the flow.

To confirm this quantitatively, I have averaged B_r along the line marked in Fig. 1. Figure 3 plots this quantity against time in log-linear coordinates; we observe oscillations and exponential growth at a rate comparable with growth rates in ref. 8. This shows that there is constructive folding taking place in the chaotic regions of the flow and indicates fast dynamo action, as the main effect of weak diffusion is to smear out fine structure in the field^{10–13} seen in Fig. 2, but to leave behind the average field, which is growing on the convective timescale (Fig. 3).

In conclusion, we observe stretching and constructive folding of magnetic field in chaotic regions of an *ABC* flow. This provides evidence for fast dynamo action in a smooth, steady, chaotic flow, although further research is needed to understand the role of the very weak diffusion. The theoretical study indicates fast dynamo processes that can occur in the steady flow of a highly conducting fluid and which may underlie the fast solar dynamo. □

Received 5 November 1990; accepted 21 February 1991.

1. Vainshtein, S. I. & Zeldovich, Ya. B. *Usp. Fiz. (SSSR)* **106**, 431–457 (1972); (Engl. transl.) *Sov. Phys. Usp.* **15**, 159–172 (1972).

entropy, which is the sum of the positive Liapunov exponents (see, for example, ref. 7), is related to the average rate of loss of information from a time series, and gives a measure of how chaotic a system is. The fractal dimension is the dimensionality of the phase space explored by the trajectory⁷, and provides an estimate of the number of independent degrees of freedom in the underlying motion. The most commonly used numerical approach⁸ yields lower bounds to the fractal dimension and the Kolmogorov entropy, thus providing a sufficient condition for chaos. But difficulties arise in estimating the fractal dimension from series with too few points⁹, and more-sophisticated treatments are necessary¹⁰. An alternative viewpoint in terms of the K-entropy would be very helpful.

Here I use the fact that the K-entropy represents the average rate of loss of information from the time series, and employ a fundamental result of information theory¹¹. The key parameter in the present analysis is the correlation coefficient between the forecast values of the time series and those actually observed. The initial decay of this coefficient as forecasts are made further into the future can be related to the average rate of loss of information. The forecasting method itself need not concern us here because the theory is independent of how the forecasts are made; results accurate to better than 0.01% are obtained for model tent-map series (see later) of 1,000 data points.

Suppose that we have an algorithm for predicting the future terms in a time series from any given point. Let the predicted values be denoted by $\{x_i\}$ and the corresponding observed values by $\{y_i\}$. Pearson's correlation coefficient between these two distributions, r , ranges in magnitude between zero and one for uncorrelated and identical distributions respectively. We are concerned with the decay of this coefficient as the forecasts, $\{x_i\}$, are made further into the future. To demonstrate the relationship between the correlation coefficient and the information entropy, we consider the statistics of the distribution defined by the forecasting errors, $\{x_i - y_i\}$. If $\langle x_i \rangle = \langle y_i \rangle$, where the angle brackets denote the average over all values of the index i , then the variance of this distribution is given by

$$\sigma^2(x_i - y_i) = \langle x_i^2 \rangle + \langle y_i^2 \rangle - 2\langle x_i y_i \rangle \quad (1)$$

so that

$$r = \frac{-\frac{1}{2}\sigma^2(x_i - y_i) + \frac{1}{2}(\langle x_i^2 \rangle + \langle y_i^2 \rangle) - \langle x_i \rangle \langle y_i \rangle}{\sigma(x_i)\sigma(y_i)} \quad (2)$$

If the first and second moments of the distribution $\{x_i\}$, in other words the mean and variance, are the same as those of the $\{y_i\}$ distribution, then

$$r = 1 - \frac{1}{2} \frac{\sigma^2(x_i - y_i)}{\sigma^2(y_i)} \quad (3)$$

The information content, I , of a distribution can be expressed in terms of the probability distribution function using expressions first derived¹¹ in connection with code-breaking. For a continuous probability distribution

$$I = \int \rho \ln \rho \, d\Delta \quad (4)$$

where $\rho(\Delta) \, d\Delta$ is the probability of the observation lying between Δ and $\Delta + d\Delta$, and in this case the specific values may be labelled as $\Delta_i = x_i - y_i$. The time derivative of I is simply $\dot{I} = -K$ where K is the Kolmogorov entropy, equivalent to the mean loss of information per unit time. Here the forecasting time in question measures the distance into the future for which the forecasts are being made for each of the data points. It is convenient to measure this forecasting time in terms of the separation between data points of the original series, which is assumed to be constant. As the present analysis considers forecasting in only one dimension, it actually yields the largest

Liapunov exponent. The formulae for K will therefore be correct only for systems with at most one positive exponent, and otherwise provide a lower bound to the true K -entropy. To relate the rate of loss of information to the statistics of $\{\Delta_i\}$, we must assume a form for $\rho(\Delta)$, and in view of the central limit theorem, a sensible choice is the normal distribution

$$\rho = \frac{1}{\sqrt{2\pi}s} e^{-\Delta^2/2s^2} \quad (5)$$

where the mean of the distribution is assumed to be zero for short-time forecasts and the variance is s^2 . Thus we obtain

$$\begin{aligned} \dot{I} = -K &= \int \dot{\rho}(1 + \ln \rho) \, d\Delta \\ &= \dot{s} \int \rho \left(\frac{\Delta^2}{s^3} - \frac{1}{s} \right) (1 + \ln \rho) \, d\Delta \\ &= \dot{s} \left[\left(\frac{\langle \Delta^2 \rangle}{s^3} - \frac{1}{s} \right) (1 - \ln \sqrt{2\pi}s) - \frac{\langle \Delta^4 \rangle}{2s^5} + \frac{\langle \Delta^2 \rangle}{2s^3} \right] \\ &= \dot{s} \left(\frac{1}{2s} - \frac{\langle \Delta^4 \rangle}{2s^5} \right) = -\frac{\dot{s}}{s} \end{aligned} \quad (6)$$

because $\langle \Delta^2 \rangle = s^2$ and $\langle \Delta^4 \rangle = 3s^4$. This result is independent of the mean and its possible dependence upon the forecasting time. Hence we arrive at the solution $s^2(t) = s^2(0) e^{2Kt}$, and therefore

$$r = 1 - \frac{1}{2} \frac{s^2(0) e^{2Kt}}{\sigma^2(y_i)}$$

or

$$\ln(1-r) = \ln[s^2(0)/2\sigma^2(y_i)] + 2Kt \quad (7)$$

We can obtain K and $s^2(0)$ from the initial slope and intercept of a plot of $\ln(1-r)$ against the forecasting time. $s^2(0)$ is a measure of the amount of information initially present and therefore of the quality of the database. We can also find a form for the decay of the information entropy; the result is

$$I(t) = -\ln(\sqrt{2\pi}e s(0)) - Kt \quad (8)$$

and so the initial information is identified as $I(0) = -\ln(\sqrt{2\pi}e s(0))$ which is large and positive for small $s(0)$. The divergent nature of this solution for $I(t)$ is characteristic of continuous distributions (see, for example, ref. 12).

The above approach provides an interesting contrast to existing methods for the systematic calculation of Liapunov exponents that have been applied to much longer time series^{13,14}. In both cases we consider the time evolution of points that are initially nearest neighbours. In the existing methods, however, these data are used to reconstruct matrices that describe the local linearized dynamics of the system. Here, the time evolution is followed over a number of steps, and the data are used to make forecasts. As shown above, the largest Liapunov exponent can be calculated from the decay in accuracy of these forecasts.

To test the theory, I have made a series of forecasts for the iterative tent map (for example, see ref. 15) $x_{n+1} = z(1 - 2|1/2 - x_n|)$, which is chaotic for $z > 1/2$ with Liapunov exponent $\ln 2z$. The forecasts used a method similar to that of Sugihara and May³. Figure 1a shows typical decay curves for r as a function of the number of iterations into the future for which forecasts were made. In this case (with $z=1$) the series consisted of 1,000 points, of which the first 500 were used as a database for making predictions about the future behaviour of the last 500. For example, to calculate the correlation coefficient for predictions one time step into the future, points x_1 to x_{500} were used to predict the values obtained by applying the iterative map to points x_{501} to x_{999} , and the results were compared with the actual values x_{502} to $x_{1,000}$. Using just the first two points on the curve in Fig. 1a and the relation in equation (7) gives $K = 0.69310$,

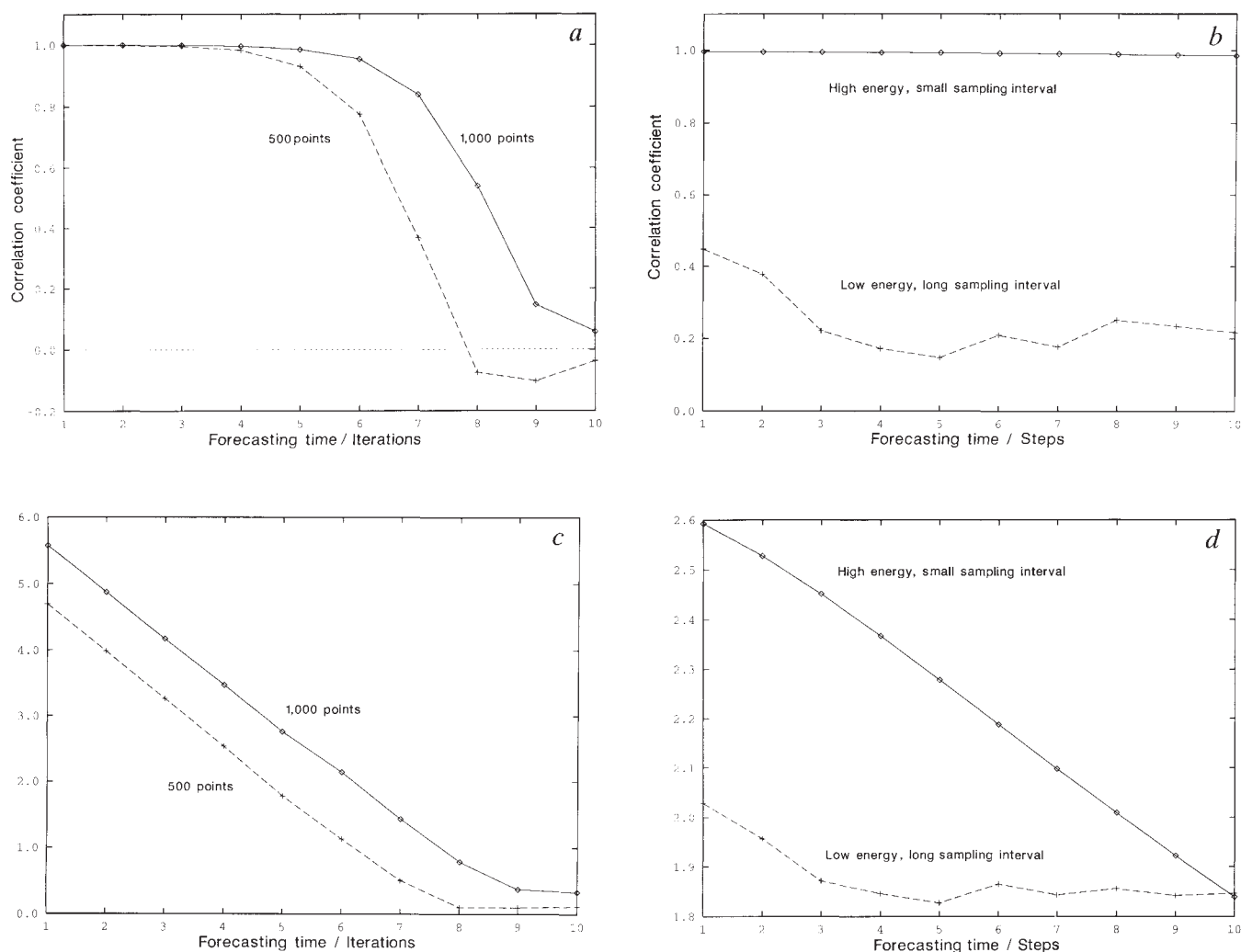


FIG. 1 *a*, Decay of the correlation coefficients with forecasting time for series of 1,000 and 500 tent map points ($z=1$). Here, and elsewhere, the first half of the series is used as the database for predicting the future behaviour of the points in the remaining half. *b*, Decay of the correlation coefficients obtained for two Ar_3 time series each consisting of 1,000 points. The upper curve is for a high energy and a short sampling interval, the lower

curve for a low energy and a long sampling interval. *c*, Results for the two tent-map series defined in *a*. The vertical axis corresponds to the natural log of the function on the right-hand side of equation (9), and the horizontal axis is the forecasting time. *d*, Results for the two Ar_3 time series defined in *b*. Axes as in *c*.

compared with the value of $\ln 2 = 0.69315 \dots$, and $s^2(0) = 3.00 \times 10^{-5}$. Fitting to the first four points gives $K = 0.6897$ and $s^2(0) = 3.03 \times 10^{-5}$ (Fig. 1c). If we split and analyse a 500-point series in the same manner then the corresponding numbers are $K = 0.706$ and $s^2(0) = 1.41 \times 10^{-4}$ for the first two points only, and $K = 0.684$, $s^2(0) = 1.51 \times 10^{-4}$ for the first four points. The initial variance is larger for the shorter series, as expected, and for both the 1,000-point and the 500-point series the results for the first-difference series are very similar to the parent series. Furthermore, a better approximation is possible in which the variance has the correct behaviour for long forecasting times and we do not assume that the first or second moments of the distributions $\{x_i\}$ and $\{y_i\}$ are the same. The first refinement replaces s by $s(1-s/\alpha)^{-1}$, where $\alpha^2 = 2\langle y_i^2 \rangle$ is the expected variance for long forecasting times when the forecasts and observations become uncorrelated. The result is

$$ze^{-Kt} = \sqrt{\frac{2\langle y_i^2 \rangle}{\langle x_i^2 \rangle + \langle y_i^2 \rangle - 2\langle x_i \rangle \langle y_i \rangle - 2r\sigma(x_i)\sigma(y_i)}} - 1 \quad (9)$$

where $z = \alpha/s(0) - 1$, which gives significantly better fits when the longer-time forecasts are included.

To test the performance of the model for real systems, I

applied the method to the motion of an Ar_3 cluster as modelled by molecular dynamics. This system is appropriate because the K-entropy has been determined independently (D.J.W. and R. S. Berry, unpublished work, and ref. 16), and results can be obtained for a range of energies and for different sampling intervals. Systematic changes in K-entropy and fractal dimension have previously been related to the 'melting' of this weakly bound atomic complex¹⁶. The K-entropy of such systems is of great importance in understanding intramolecular energy transfer between different degrees of freedom, which may be investigated spectroscopically in vibrationally excited molecules, for example. The larger the K-entropy, the more chaotic the dynamics and the faster the energy is redistributed. Only if specific excitations are rapidly relaxed throughout all the molecular degrees of freedom are we justified in applying standard statistical theories of chemical reactions.

The x coordinate of one of the argon atoms was used to construct time series of 1,000 points which were sampled at intervals of 2–100 time-steps of length 1×10^{-14} s. Because there are two positive Liapunov exponents, the present approach should give a lower bound to the K-entropy, as explained above. There are many more degrees of freedom than in the tent map, as well as a certain amount of regularity due to vibrational

motion. Furthermore, the cluster dynamics correspond to a nonlinear map; we may judge the likely effect of this from the logistic map $x_{n+1} = 4x_n(1-x_n)$, which has the same Liapunov exponent¹⁵ ($\ln 2$) as the tent map with $z = 1$ and may be transformed into it. Overall fits for the logistic map were not quite as good as for the tent map; the best estimate of K from 1,000 points was 0.72.

Tests were performed for the cluster at a range of energies up to the evaporation limit; the plots of $\ln(1-r)$ against forecasting time generally exhibited slopes that decreased in magnitude with increasing forecasting time, and estimates of K from the initial slopes decreased as the sampling interval between points was increased. Order-of-magnitude estimates of the K-entropy from these initial slopes, however, certainly exhibit the correct trend as the energy of the trajectory is changed (D.J.W. and R. S. Berry, unpublished work; and ref. 16): there is a distinct difference between low-energy series, for which the K-entropy is zero, and high-energy series, for which K values of $\sim 10^{12} \text{ s}^{-1}$ are estimated. At higher energies some satisfactory straight-line fits were obtained, which gave estimates of K in good agreement with the previous calculations. At lower energies, where the Ar_3 cluster is characterized by large-amplitude vibrations within a potential well, the forecasting algorithm was less effective, and spuriously high estimates of the K-entropy were obtained. Some results for Ar_3 are shown in Fig. 1b and d.

Time series constructed from a sine function plus a tent-map sequence show similar effects: the initial slopes of the $\ln(1-r)$ curves give values for the K-entropy that are too small and decrease as the frequency of the regular sine-wave component decreases. The greater the amplitude of the regular part, the poorer are the estimates of the K-entropy (usually smaller than the known values). The estimates improve if the periodic component of the signal is removed by determining the largest components of the corresponding Fourier series and subtracting them. It is important not to subtract a more accurate Fourier series representation because this would result in a completely flat signal. When the ratio of the chaotic signal to the regular component is boosted in this way, the estimates of the K-entropy from the initial slope become fairly accurate for lower sine-wave frequencies. Partial removal of the regular components of the Ar_3 time series, however, did not improve the results significantly.

The encouraging results obtained for short time series illustrate how this approach provides a measure of the degree of chaos in the underlying dynamics. The theory is independent of the forecasting method, which might be improved in future work¹⁷ along with the treatment of series containing significant regular components. Analyses of some geological sparse time series will be reported in a future communication (G. P. Weedon and D.J.W., unpublished work); the present method should at least tell us whether such sequences contain sufficient data points to be useful, and may enable us to quantify just how chaotic the underlying dynamics are. $\square$

High-pressure dielectric measurements of solid hydrogen to 170 GPa

R. J. Hemley, M. Hanfland & H. K. Mao

Geophysical Laboratory, Carnegie Institution of Washington, 5251 Broad Branch Rd, N.W., Washington, DC 20015, USA

At high pressure ($>100 \text{ GPa}$), the valence-conduction band gap in solid hydrogen is predicted to decrease and eventually to close, transforming it from an insulator to a metal¹. Recent experiments at pressures up to $\sim 300 \text{ GPa}$ suggest a gradual closing of the direct gap at very high pressures whereas closure of the indirect gap may occur at lower pressures, perhaps below 200 GPa (refs 2–6). Measurements of the refractive index, n , and its frequency dependence (the dispersion $dn/d\omega$) as a function of pressure can provide information on the changes in electronic structure that occur under these conditions^{7,8}. Here we report refractive-index measurements on solid hydrogen at visible frequencies at pressures up to 170 GPa . Unlike earlier studies⁹, we find no evidence for a divergence (dielectric instability) at 150 GPa , close to the low-temperature phase transition observed previously⁶ and suggested as being associated with metallization^{4,9}. Our results are consistent with closure of the indirect gap. A fit of our data to a dielectric model indicates that the onset of visible absorption owing to direct interband transitions should occur above 200 GPa , consistent with previous direct observations³. Pressure-induced molecular dissociation may occur before closure of the direct gap.

At zero pressure the lowest-energy electronic transitions in solid hydrogen consist of excitonic bands at 11 eV , with excitation from the valence to the conduction band appearing at 15 eV (ref. 10). Recent experimental studies³ using diamond-cell techniques have shown evidence for direct electronic transitions in the molecular solid at visible wavelengths for pressures above 200 GPa . Increases in infrared reflectivity and absorption^{4,5} and the occurrence of a low-temperature phase transition⁶ near 150 GPa have been reported. Measurements of the optical properties of samples at static high pressures are limited, however, to photon energies below $3\text{--}4 \text{ eV}$ by the diamond window of the cell (type Ia diamonds). This limitation precludes direct measurement of valence-conduction band transition energies of hydrogen by one-photon absorption at lower pressures. Accurate measurements of the index of refraction for hydrogen have been obtained up to 20 GPa in a diamond-cell by Brillouin scattering spectroscopy¹¹. Alternatively, the index of refraction of samples at high pressure can be measured interferometrically. If the diamond flats of the compression cell remain parallel, and the index of refraction of the sample differs sufficiently from that of diamond, Fabry-Perot interference fringes can be measured under very high loads, as was originally demonstrated for solid argon¹². Van Straaten and Silvera¹³ have determined n and $dn/d\omega$ for hydrogen at pressures up to 37 GPa by the use of this interference technique, and they fitted their data to a dielectric model. Recently, Eggert *et al.*⁹ have carried out a similar study, obtaining two data points (at 67 and 73 GPa). A linear extrapolation of these results to higher pressures suggests that a dielectric catastrophe ($n \rightarrow \infty$) occurs at $148 (\pm 19) \text{ GPa}$ (ref. 9).

We loaded hydrogen in a modified megabar-type diamond cell as described previously^{4,5}. The diameters of the samples were $20\text{--}30 \mu\text{m}$, so that the sample volumes were about an order of magnitude larger than those used in the previous optical studies³ at $200\text{--}300 \text{ GPa}$. The larger sample size means that the maximum pressures attainable are lower¹⁴, but it significantly improves the accuracy with which the optical constants can be measured. Interference fringes were measured in reflectance mode with both diode array and scanning spectrometer sys-

Received 21 November 1990; accepted 28 February 1991.

- Olsen, L. F. & Schaffer, W. M. *Science* **249**, 499–504 (1990).
- Nicolis, C. & Nicolis, G. *Nature* **311**, 529–532 (1984).
- Sugihara, G. & May, M. *Nature* **344**, 734–741 (1990).
- Farmer, J. D. & Sidorowich, J. J. *Phys. Rev. Lett.* **59**, 845–848 (1987).
- Kolmogorov, A. N. *Dokl. Akad. SSSR* **124**, 754 (1959).
- Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1983).
- Schuster, H. G. *Deterministic Chaos* 2nd ed. (VCH, Weinheim, 1989).
- Grassberger, P. & Procaccia, I., *Phys. Rev.* **A28**, 2591–2593 (1983).
- Grassberger, P. *Nature* **323**, 609–612 (1986).
- Vautard, R. & Ghil, M. *Physica* **D35**, 395–424 (1989).
- Shannon, C. & Weaver, W. *The Mathematical Theory of Communication* (Illinois University Press, Urbana, 1949).
- Schlögl, F. *Probability and Heat* (Vieweg, Braunschweig, 1989).
- Sano, M. & Sawada, Y. *Phys. Rev. Lett.* **55**, 1082–1085 (1985).
- Eckmann, J.-P., Kamphorst, S. O., Ruelle, D. & Ciliberto, S. *Phys. Rev.* **A34**, 4971–4979 (1986).
- Lichtenburg, A. J. & Leiberman, M. A. *Regular and Stochastic Motion* (Springer, New York, 1983).
- Beck, T. L., Leitner, D. M. & Berry, R. S. *J. chem. Phys.* **89**, 1681–1694 (1988).
- Casdagli, M. *Physica* **D35**, 335–356 (1989).

ACKNOWLEDGEMENTS. The author is a Lloyd's Tercentenary Fellow and a Fellow of Downing College, Cambridge. Some of the computer time for this project was furnished by a grant from the SERC.

tems^{3,4}. Pressures were determined by ruby R₁ fluorescence from small grains of ruby¹⁵, which in each case were estimated to occupy less than ~5% of the total sample area, and by Raman measurements of the vibron frequency, which has been calibrated against the quasihydrostatic ruby R₁ pressure scale². Because of the possibility of loss of sample at very high pressures, the vibron was measured by Raman scattering at each pressure to ensure that hydrogen remained in the sample chamber.

Interference patterns of hydrogen at high pressures are shown in Fig. 1. The occurrence of the patterns indicates that the diamond faces remain sufficiently parallel and flat for the sample chamber of the diamond cell to function like a Fabry-Perot interferometer even at these high pressures: anvil 'cupping' effects and pressure gradients are small¹⁴. Below 110 GPa, the amplitude of the fringes decreases with increasing pressure, whereas above 140 GPa the amplitude increases. For a Fabry-Perot interferometer, the ratio between the reflected and incident light (I_R/I_0) as a function of wavelength λ is given by¹⁶

$$I_R(\lambda)/I_0(\lambda) = 1 - 1/[1 + F \sin^2(2\pi\delta/\lambda)] \quad (1)$$

where F , a measure of the amplitude of the fringes, is defined as $F = 4R/(1-R)^2$. Here R is the reflectivity, given by the Fresnel equation $R = (n_d - n_H)^2/(n_d + n_H)^2$, and $\delta = 2n_H d \cos \theta$ is the optical thickness of the sample and determines the wavelength interval between fringes; d and θ are the sample thickness and the angle of the incident light; n_d and n_H are refractive indices of diamond and hydrogen; and $\cos \theta \approx 1$ for our measurements (near normal incidence). The index of refraction of the sample, n_H , can be readily determined from the distance between the fringes or the positions of the maxima and minima if d is known. Alternatively, it can be determined from the amplitude of the fringes if the value of n_d is known.

Here we focus on the pressure variation of the dispersion of the hydrogen refractive index, $dn_H/d\omega$. Wemple and DiDomenico¹⁷ have shown that the dispersion in refractive index for a wide range of substances below the valence-conduction band edge can be approximated by an expression involving a single Sellmeier oscillator,

$$n(\omega)^2 - 1 = F_1/(\omega_1^2 - \omega^2) \quad (2)$$

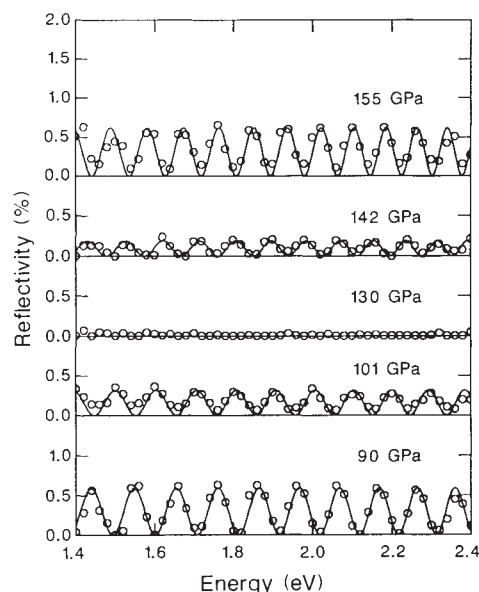
where F_1 is the oscillator strength and ω_1 is an effective oscillator frequency, which is correlated closely with the energies of direct valence-conduction band transitions¹⁷. We determined the dispersion $dn_H/d\omega$ at each pressure from the positions of the maxima and minima of the interference pattern. This analysis

requires an accurate assignment of the order of each fringe. To solve this problem, we made two runs with different sample thicknesses (varying between 1 and 3 μm) and performed measurements at several identical pressures. The dispersion was calculated for various assignments of the order of the fringes, and the correct order assignment was that for which the value of the dispersion matched in the two runs. We fitted a combination of equations (1) and (2) to our experimental spectra by a least-squares fit. The calculated interference patterns are in very good agreement with the measurements (Fig. 1).

We find that the oscillator frequencies ω_1 decrease with increasing pressure, which is consistent with the low-pressure results^{9,13}. The behaviour of ω_1 is close to linear in density (rather than pressure or volume). The values for ω_1 are plotted in Fig. 2a as a function of relative density ρ/ρ_0 using the experimentally determined equation of state¹⁸. The results are consistent with those obtained previously at lower pressure using the same model^{9,13}. We also include the previous extrapolated model calculations based on the low-pressure measurements⁹ ($P < 73$ GPa; $\rho/\rho_0 < 7$). The uncertainty in these extrapolations demonstrates the necessity of direct measurements at higher pressures: the linear pressure extrapolation⁹ predicts a dielectric catastrophe at $\rho/\rho_0 = 9$ and a value of n_H of ~100 at 125 GPa, whereas our direct observations (Fig. 1) indicate no anomalous behaviour in this pressure range ($n_H \leq 3.0$ at 125 GPa)⁴.

We carried out a number of measurements as a function of pressure through the low-temperature phase transition that occurs⁶ at 150 GPa (77 K). This transition is characterized by a marked discontinuity in the intramolecular vibrational frequency (vibron)^{6,19,20}; the most recent measurements indicate that it is primarily an electronic change associated with bond weakening²¹. Measurements of the temperature dependence of the transition²² provide evidence for a critical point along the phase line, such that the transition is continuous above ~150 K. The observation of increases in infrared reflectivity⁴ and absorption⁵ in the pressure range of the transition suggest that it is associated with metallization. We find that the interference pattern measured in the visible varies continuously through the transition. The change in n across the transition is less than 5%, which in turn implies that the density change associated with the transition may be of similar magnitude. The observation of fringes confirms that the absorption is negligible in this spectral range^{5,23} and further demonstrates that if metallization occurs by band overlap at these pressures, the gap closure must be indirect (that is, overlapping parts of the bands are in different

FIG. 1 Representative interference patterns of hydrogen at high pressure and 295 K. The solid lines correspond to simultaneous fits of equations (1) and (2) to the measured data (open circles). A detailed study of the pressure dependence of n_H will be presented elsewhere.



regions of k -space): metallization need not be associated with a change in optical properties at visible wavelengths if gap closure is indirect and if the plasma frequency remains in the infrared range^{4,5}.

Wemple and DiDomenico¹⁷ have shown that the frequency of the single oscillator in equation (2) represents the energy of an average gap, rather than that of specific indirect or direct gaps. The value of ω_1 at zero pressure ($\rho/\rho_0 = 1$) determined from gas-phase data²⁴ (corrected for the density change) is 13.5 eV, which is between the strong bands observed experimentally at 11 and 15 eV (Fig. 2b)¹⁰. Up to the maximum pressure of our investigation (170 GPa), the best-fit value of ω_1 (4.8 eV) remains in the ultraviolet range, consistent with observations that hydrogen samples remain transparent to these pressures. If this correlation extends to higher pressure and ω_1

remains linear in density, our analysis indicates that visible absorption should be apparent at $\rho/\rho_0 > 10$ –11 (above 200 GPa). This result is in accord with the optical evidence³ for darkening of the sample at 250–300 GPa. Extrapolating further, we find that $\omega_1 = 0$ at $\rho/\rho_0 \approx 14$. Although closure of the direct band gap of hydrogen might therefore be expected in this higher density range, other phase transitions, including molecular dissociation, are likely to intervene³. The present study underscores the need for direct measurements at ultrahigh pressures, rather than relying on the results of extrapolation, to examine this question.

Several electronic-structure calculations of the pressure dependence of the band gap of solid molecular hydrogen have been performed^{1,25–27}. Detailed experimental tests of the theoretically predictions, however, have not been possible. It is therefore instructive to compare the measured density dependence of ω_1 with the band gaps predicted theoretically. In Fig. 2b we show the experimental ω_1 and theoretical results calculated for different high-pressure crystal structures (cubic $Pa3$ ^{2,25} and both ordered and disordered hexagonal close-packed^{26,27}). We find that the theoretical calculations all show linear behaviour over an extended density range, in agreement with the observed behaviour of ω_1 . Calculations based on the local-density approximation (LDA)^{1,25,26} all underestimate the band gap at ρ_0 , whereas the value derived in the first-principles quasiparticle study²⁷ is in good agreement with the experimental result¹⁰. Despite this deficiency, the slopes of several theoretical curves are comparable with the experimental result $d\omega_1/d(\rho/\rho_0) = 0.96$ eV. The underestimation of the band gap by the LDA was corrected in an *ad hoc* fashion by Garcia *et al.*²⁶ by use of the $X\alpha$ approximation for the exchange potential. The agreement with experiment for this calculation, assuming a disordered h.c.p. structure, is close. The calculated result, however, is for the indirect gap; it would be useful to examine whether or not theoretical predictions for the average energy gap follow a similar trend. The spectroscopic data suggest that the underlying lattice is derived from h.c.p.²¹, although detailed information on the crystal structure above 40 GPa (from X-ray diffraction, for example) is not available at present²⁸. Garcia *et al.*²⁶ predict that there is a significant difference in electronic properties for h.c.p. hydrogen in c -axis-aligned and disordered structures, and they further suggest that the 150-GPa transition is an orientational ordering transition involving this change. The lack of large changes in the dielectric properties across the transition found here suggests that the transition is not of this type, a conclusion supported by the character of the low-frequency vibrational spectrum²¹. Our present results thus provide important new constraints concerning the nature of the phase of hydrogen formed at low temperature above 150 GPa. □

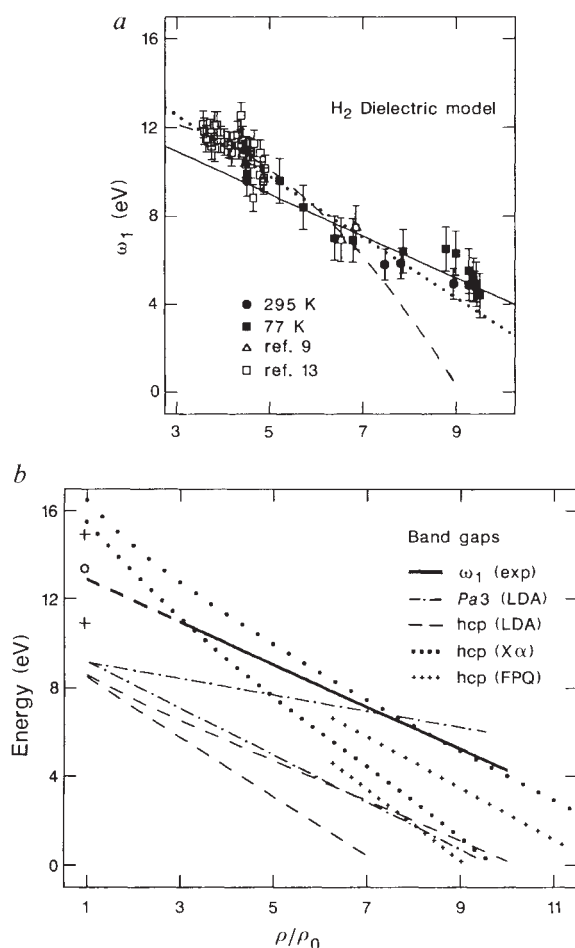


FIG. 2 *a*, Effective oscillator frequency of hydrogen as a function of relative density. The solid line is the best fit to the present data. The dashed line is the result of Egger *et al.*⁹ extrapolated above 73 GPa ($\rho/\rho_0 = 6.9$) assuming a linear dependence on pressure. The dotted line is the extrapolated result assuming a linear dependence on density based on our subsequent analysis of theoretical results (see addendum in ref. 9). Direct measurements are required to choose between the two extrapolations. *b*, Comparison of density dependence of the effective oscillator frequency and theoretical predictions of the direct and indirect band gap. The experimentally determined exciton and band-gap energies at $\rho/\rho_0 = 1$ are indicated by crosses¹⁰. The value of ω_1 determined from the gas-phase data, corrected for the density of the solid at zero-pressure using the Lorentz relation⁸, is indicated by the circle. The upper and lower curves calculated for the $Pa3$ structure are the direct and indirect gaps, respectively, determined by Friedli and Ashcroft¹ and Min *et al.*²⁵. The upper and lower curves calculated for the h.c.p. structure are those derived by Garcia *et al.*²⁶ (using the LDA, $X\alpha$ formalism) and Chacham and Louie²⁷ (FPQ) assuming a spherical (rotationally averaged) potential and molecular ordering parallel to the c axis, respectively.

Received 11 January; accepted 1 March 1991.

- Friedli, C. & Ashcroft, N. W. *Phys. Rev. B* **16**, 662–672 (1977).
- Hemley, R. J., Mao, H. K. & Hanfland, M. in *Molecular Systems Under High Pressure* (eds Pucci, R. & Piccolto, G.) 223–243 (North-Holland, Amsterdam, 1991).
- Mao, H. K. & Hemley, R. J. *Science* **244**, 1462–1465 (1989).
- Mao, H. K., Hemley, R. J. & Hanfland, M. *Phys. Rev. Lett.* **65**, 484–487 (1990).
- Hanfland, M., Hemley, R. J. & Mao, H. K. *Phys. Rev. B* (in the press).
- Hemley, R. J. & Mao, H. K. *Phys. Rev. Lett.* **61**, 857–860 (1988).
- Herzfeld, K. *Phys. Rev.* **29**, 701–705 (1927).
- Wooten, F. *Optical Properties of Solids* (Academic, New York, 1972).
- Egger, J., Goettel, K. A. & Silvera, I. F. *Europhys. Lett.* **11**, 775–781 (1990); addendum **12**, 381–382 (1990).
- Inoue, K., Kanzaki, H. & Suga, S. *Solid St. Commun.* **30**, 627–629 (1979).
- Shimizu, H., Brody, E. M., Mao, H. K. & Bell, P. M. *Phys. Rev. Lett.* **47**, 128–131 (1981).
- Grimsditch, M., Letoullec, R., Polian, A. & Gauthier, M. *J. appl. Phys.* **60**, 3479–3481 (1986).
- Van Straaten, J. & Silvera, I. F. *Phys. Rev. B* **37**, 6478–6481 (1988).
- Mao, H. K. in *Simple Molecular Systems at Very High Densities* (eds Loubeyre, P., Polian, A. & Boccaro, N.) 221–236 (Plenum, New York, 1989).
- Mao, H. K., Xu, J. & Bell, P. M. *J. geophys. Res.* **91**, 4673–4676 (1986).
- Born, M. & Wolf, E. *Principles of Optics* (Pergamon, Oxford, 1980).
- Wemple, S. H. & DiDomenico, M. *Phys. Rev. B* **3**, 1338–1351 (1971).
- Mao, H. K. *et al. Science* **239**, 1131–1134 (1988).
- Hemley, R. J. & Mao, H. K. *Phys. Rev. Lett.* **63**, 1393–1395 (1989).
- Lorenzana, H. E., Silvera, I. F. & Goettel, K. A. *Phys. Rev. Lett.* **63**, 2080–2083 (1989).
- Hemley, R. J., Mao, H. K. & Shu, J. F. *Phys. Rev. Lett.* **65**, 2670–2673 (1990).
- Hemley, R. J. & Mao, H. K. *Science* **249**, 391–393 (1990).

23. Eggert, J. H. *et al.* *Phys. Rev. Lett.* **66**, 193–196 (1991).
 24. Peck, E. R. & Huang, S. J. *Opt. Soc. Am.* **67**, 1550–1554 (1977).
 25. Min, B. I., Jansen, H. J. F. & Freeman, A. J. *Phys. Rev. B* **33**, 6383–6390 (1986).
 26. Garcia, A., Barbee, T. W., Cohen, M. L. & Silvera, I. F. *Europhys. Lett.* **13**, 355–360 (1990).
 27. Chacham, H. & Louie, S. G. *Phys. Rev. Lett.* **66**, 64–67 (1991).
 28. Hemley, R. J. *et al.* *Eos* **71**, 1620 (1990).

ACKNOWLEDGEMENTS. We thank J. Shu for help with the experiments. This work was supported by the NSF, NASA and the Carnegie Institution of Washington.

Depletion of barium and radium-226 in Black Sea surface waters over the past thirty years

K. Kenison Falkner^{*‡}, D. J. O'Neill^{†‡}, J. F. Todd^{†‡},
W. S. Moore[†] & J. M. Edmond^{*}

^{*} Earth, Atmospheric and Planetary Sciences,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

[†] Geological Sciences, University of South Carolina, Columbia,
South Carolina 29208, USA

THE nearly landlocked waters of the Black Sea support a valuable fishery, but are also particularly vulnerable to anthropogenic disturbance. Here we use dissolved barium and radium-226 as tracers, to investigate the biogeochemical health of the sea. Both elements are brought to surface waters by vertical mixing of deeper, enriched waters^{1,2}, and by rivers^{3–8}; these inputs should ordinarily be balanced by outflow of surface waters at the Bosphorus, and by biologically mediated removal of ²²⁶Ra-bearing barite. We show, however, that surface-water inventories have been substantially depleted over the past few decades: recent (1988–89) barium concentrations were 1.6 times lower than in 1975⁹; and recent ²²⁶Ra concentrations were 3.5 times lower than in 1958¹⁰ and 1967¹¹. These observations suggest that steady-state cycling of these elements has been perturbed by increased primary productivity, presumably fuelled by nutrients from industry and agricultural runoff, and to a lesser extent by decreased fluvial sediment loads owing to extensive impoundment of rivers in the region.

Samples were collected during the following cruises: RV *Bilim* (April 1990); RV *Vodjanitsky* N28 (April 1989); RV *Knorr* Black Sea Expedition, legs 3, 4 and 6 (June 1988, June–July 1988 and August 1988, respectively); and an RV *Akademik A. Kovalevskii* cruise (August 1990). Station locations are shown in Fig. 1. Surface samples were taken from depths of 5–10 m within the mixed layer. Appropriate precautions were taken to avoid contamination and ensure sample viability during handling and storage^{2,12}. Barium analyses were accomplished by isotope dilution–inductively coupled plasma quadrupole mass spectrometry (ID–ICPMS) on filtered samples (0.4 µm pore size) and unfiltered samples that were diluted 80–100 times with 0.2 M HNO₃ (refs 2, 13). After quantitative preconcentration of ~20 l unfiltered water on acrylic fibres impregnated with manganese dioxide, ²²⁶Ra was determined by ²²²Rn emanation¹⁴. Precisions based on replicate analyses are 3% for Ba and 10% for ²²⁶Ra. Table 1 summarizes our results, together with previously reported values.

Within the 3% precision of the Ba determinations, there were no concentration differences between filtered and unfiltered samples, indicating that Ba is present primarily in dissolved form. The same is probably true for ²²⁶Ra. Concentrations of

both Ba and ²²⁶Ra were fairly homogeneous over the broad areas sampled in 1988–1990, averaging 195 ± 16 nM and 6.6 ± 1.1 d.p.m. per 100 l respectively. Lower Ba concentrations (~140 nM) were obtained in 1990 at coastal locations which may not be representative of the surface Black Sea. Standard deviations of the averages only slightly exceed analytical precision, and spatial variability appears to be insignificant with respect to the differences of 115 nM and 15 d.p.m. per 100 l between present-day values and ones reported 15–30 years ago. Although there is only one previous Ba concentration measurement available, the facts that all previous ²²⁶Ra concentrations are higher than ours and that samples were taken at comparable times of the year suggest a temporally decreasing trend. Winter months are not represented in Table 1, but it seems unlikely that seasonal variability could produce concentration changes as large as those reported, as the rapid surface currents of the Black Sea, averaging 30–40 cm s⁻¹ (refs 15, 16), tend to homogenize surface properties in a few months. This should be verified by seasonal studies of Ba and ²²⁶Ra on a broad spatial scale. The concentration differences cannot be attributed to seasonality in riverine inputs as the Danube, the dominant riverine source (K.K.F., W.S.M. and J.M.E., unpublished data), does not undergo large seasonal changes in flow or suspended sediment load¹⁵. The reported concentration differences are not attributable to analytical offsets or error; redetermination of Ba by ID–ICPMS in samples archived from the 1975 expedition resulted in the values originally found by thermal ionization mass spectrometry⁹, and open-ocean ²²⁶Ra values reported for the 1967 expedition are similar to those in recent determinations^{11,17,18}. To explain the declining levels of Ba and ²²⁶Ra, we must examine processes governing their distributions in the Black Sea.

The formation of barite in association with biological activity is the dominant removal pathway for Ba and its chemical analogue ²²⁶Ra from open-ocean surface waters^{19–22}. The mechanism of barite formation in undersaturated sea water^{2,23} has yet to be fully explained, but in the open ocean it seems to occur in microenvironments provided by the broken siliceous tests of dead diatoms and associated decaying organic matter²⁴. The surface of the Black Sea sustains large populations of diatoms which can dominate primary production there²⁵. Trapping of sediment over the years 1983–1988 at southwestern Black Sea sites showed that despite variations of several orders of magnitude, fluxes of Ba and organic carbon, normalized to aluminium concentrations, displayed similar trends²⁶ (B. Hay and S. Honjo, personal communication). This indicates that, as in the open ocean, Ba removal is biologically mediated. Portions of these trap samples that had high bulk Ba contents and maximal fluxes of either Si (predominantly diatoms of *Rhizosolenia* and *Chaetoceros* groups) or CaCO₃ (coccoliths of *Emiliania huxleyi*) were examined for the presence of barite by energy dispersive X-ray scanning electron microscopy.

All the trap samples contained copious Si, although only amorphous silicate and no recognizable diatom tests remained in those samples that contained large numbers of coccoliths. Barite was present in all samples. In suspended material from the open ocean, discrete 1-µm barite crystals predominate, with lesser amounts present as aggregates of smaller crystals^{19,21,24}. Almost all of the Black Sea trap barite occurred as such aggregates, 1–10 µm in diameter, composed of many (10–50) poorly crystalline 0.1–1.0-µm particles. The higher abundance of aggregates, relative to individual crystals, indicates that aggregates are the primary form of rapidly settling barite. Our observations do not allow us to assign barite formation to a particular organism, because diatoms may disintegrate during their collection or storage as wet aliquots. As coccolith dissolution tends to buffer pH to a higher value (~8.3), thus promoting diatom dissolution, the absence of diatoms in coccolith-barite-rich samples does not rule out a diatom–barite association.

Regardless of the exact nature of barite formation, enhanced

[‡] Present address: Groupe de Recherche Géodésie Spatiale, Centre National d'Etudes Spatiales, 18 Avenue Edouard Belin, 31055 Toulouse Cedex, France (K.K.F.); Dames and Moore, Geosciences Group, 911 Wilshire Boulevard, Los Angeles, California 90012, USA (D.J.O.N.); NOAA, Oceanic and Atmospheric Research, Office of Global Programs, SSMC-1, 1335 East-West Highway, Silver Spring, Maryland 20910, USA (J.F.T.).

TABLE 1 Barium and radium-226 in Black Sea surface waters

Expedition	Station	Date	Location	Position on map	Salinity* p.p.t.	Dissolved Ba nM	Ra-226 d.p.m. per 100 l
Soviet ¹⁰	S3	before August 1959	44° 33' N, 37° 58' E	1	—	—	23.90
	S171	before August 1959	43° 45' N, 36° 35' E	2	—	—	18.50
	S174	before August 1959	43° 28' N, 33° 40' E	3	—	—	19.50
	S175	before August 1959	42° 46' N, 30° 18' E	4	—	—	21.70
	S179	before August 1959	43° 23' N, 30° 02' E	5	—	—	21.70
	S181	before August 1959	43° 23' N, 30° 02' E	5	—	—	21.70
	S182	before August 1959	45° 53' N, 30° 07' E	6	—	—	20.60
	S192	before August 1959	45° 05' N, 32° 55' E	7	—	—	20.60
	S194	before August 1959	44° 10' N, 33° 10' E	8	—	—	21.70
RV <i>Oceanographer</i> ¹¹	U	before April 67	42° 47' N, 30° 16' E	4	—	—	24.60
RV <i>Chain</i> ⁹	C1355	before April 75	42° 50' N, 33° 00' E	9	18.371	310†	—
RV <i>Knorr</i>	K3-6	before June 88	43° 04' N, 34° 00' E	10	18.288	182	—
	K3-6	before June 88	43° 04' N, 34° 00' E	10	18.291	183	—
	K3-6	before June 88	43° 04' N, 34° 00' E	10	18.409	190	—
	K4-2	before June–July 1988	41° 23' N, 29° 05' E	11	18.025	—	5.50
	K4-3	before June–July 1988	41° 35' N, 29° 04' E	12	18.026	—	8.02
	K4-7	before June–July 1988	41° 51' N, 30° 21' E	13	18.350	—	5.84
	K4-8	before June–July 1988	42° 10' N, 30° 00' E	14	18.350	184	7.02
	K4-9	before June–July 1988	42° 52' N, 31° 18' E	15	18.288	243	5.00
	K4-10	before June–July 1988	42° 50' N, 32° 00' E	16	18.287	—	6.26
	K4-11	before June–July 1988	42° 10' N, 32° 45' E	17	18.491	—	5.43
	K4-12	before June–July 1988	42° 03' N, 32° 49' E	18	18.100	—	4.65
	K4-13	before June–July 1988	41° 58' N, 32° 54' E	19	18.100	—	6.21
	K4-14	before June–July 1988	43° 05' N, 34° 01' E	10	18.320	195	5.98
	K4-18	before June–July 1988	42° 45' N, 37° 35' E	20	18.320	234	6.23
	K4-19	before June–July 1988	41° 30' N, 40° 44' E	21	18.230	—	5.92
	K4-20	before June–July 1988	41° 10' N, 40° 36' E	22	18.100	—	6.15
	K6-1	before June–July 1988	41° 10' N, 29° 09' E	23	17.440	—	7.72
RV <i>Vodjanitsky</i>	V1	before April 1989	43° 41' N, 31° 19' E	24	18.26*	190	—
	V2	before April 1989	44° 58' N, 32° 42' E	25	22.01*	188	—
	V2	before April 1989	44° 58' N, 32° 42' E	25	22.01*	189	—
	V3	before April 1989	43° 57' N, 30° 37' E	26	18.19*	191	—
	V4	before April 1989	44° 14' N, 30° 32' E	27	18.26*	190	—
	V4	before April 1989	44° 14' N, 30° 32' E	27	18.24*	187	—
	V5	before April 1989	44° 43' N, 30° 13' E	28	17.60*	194	—
	V6	before April 1989	44° 53' N, 30° 37' E	29	17.29	188	—
	V7	before April 1989	44° 36' N, 31° 21' E	30	18.15*	186	—
	V8	before April 1989	46° 27' N, 31° 10' E	31	16.75*	193	—
RV <i>Bilim</i>	E2	before April 1990	40° 59' N, 28° 59' E	32	18.86	—	6.90
	K0	before April 1990	41° 00' N, 28° 59' E	32	17.90	—	7.40
	K1	before April 1990	41° 15' N, 29° 10' E	33	17.99	—	7.69
	K2	before April 1990	41° 25' N, 29° 15' E	34	18.09	—	7.06
	K5	before April 1990	41° 30' N, 29° 30' E	35	18.10	—	6.56
	K8	before April 1990	41° 30' N, 28° 30' E	36	18.05	—	5.61
	K9	before April 1990	41° 40' N, 28° 40' E	37	18.08	—	6.66
	K10	before April 1990	41° 45' N, 28° 45' E	38	18.06	—	6.57
	K11	before April 1990	41° 50' N, 28° 50' E	39	18.07	—	8.13
	K15	before April 1990	41° 40' N, 29° 30' E	40	18.10	—	8.78
	K17	before April 1990	41° 50' N, 29° 40' E	41	18.13	—	7.44
	K22	before April 1990	41° 15' N, 30° 00' E	42	18.03	—	5.60
	K26	before April 1990	41° 30' N, 30° 40' E	43	18.05	—	8.63
	K27	before April 1990	41° 25' N, 30° 40' E	44	17.98	—	7.23
	K28	before April 1990	41° 15' N, 30° 40' E	45	18.04	—	5.93
	K33	before April 1990	42° 20' N, 31° 00' E	46	18.29	—	4.77
RV <i>Kovalevskii</i>	KOV1	before August 1990	44° 34' N, 33° 22' E	47	—	155	—
	KOV2	before August 1990	44° 33' N, 33° 19' E	48	—	146	—
	KOV3	before August 1990	44° 32' N, 33° 17' E	49	—	139	—

* Calculated from measured chlorinity. A dash indicates that data were unavailable.

† Measured by thermal ionization mass spectrometry, confirmed by ID-ICPMS on archived sample.

net primary productivity is probably a direct cause of the decline in concentrations of Ba and ²²⁶Ra in surface water. Standing stocks of biomass in the most productive areas of the Black Sea (northwestern shelf and Azov region) were at least ten times as large in the 1970s than in the 1960s¹⁵, showing that productivity has indeed been increasing. Agriculture and industry both underwent vast expansion in the Black Sea watershed during this period, and the elevated productivity has been attributed to excess nutrients and organic matter from agricultural run-off and industrial effluents^{15,27}.

After the Second World War, extensive damming of the Don, Dnepr, Dneistr and Kuban rivers provided hydroelectric power which facilitated the expansion of economic activities in the region. Large-scale use of water for agriculture and industry¹⁵ began in the late 1960s. Net water withdrawals for irrigation and industry (7–11% of the annual freshwater input) have not been large enough to be directly responsible for the decreased Ba and ²²⁶Ra concentrations¹⁵. Trapping of sediments behind the impoundments, however, could lower the inputs of Ba and ²²⁶Ra normally desorbed from riverine clays in the estuary³⁸.

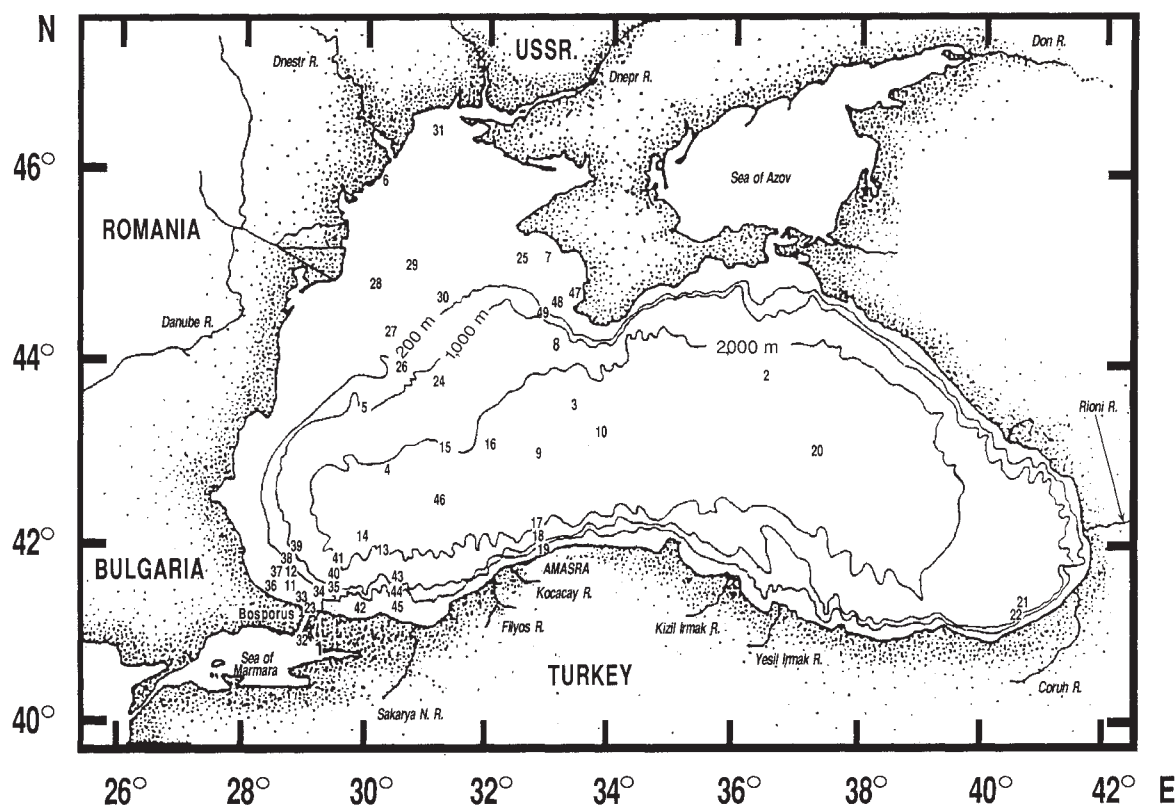


FIG. 1 Black Sea station locations. For coordinates see Table 1.

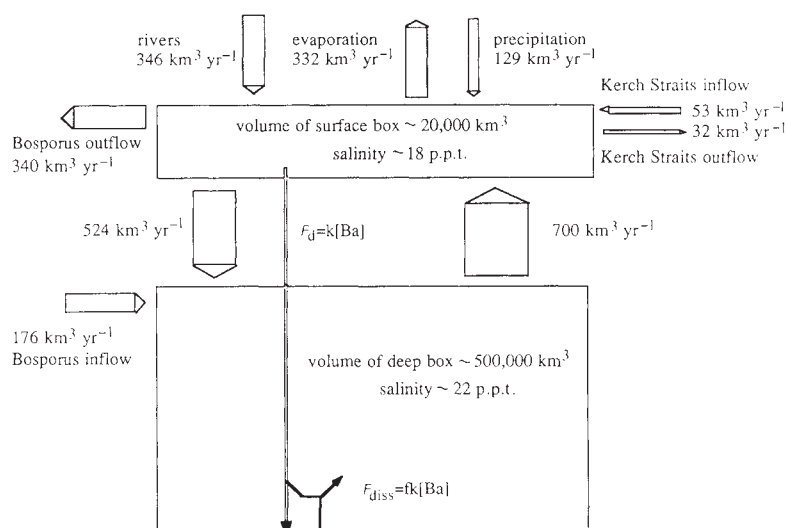
The impact of this effect is limited because the Danube has yet to have its course significantly altered. At present, canals in Romania (Danube–Constanza) and the USSR (Danube–Dniestr) divert less than 20 km^3 water and an insignificant portion of the suspended sediment load annually (G. Polikarpov, personal communication).

The change in desorbed input is a matter for speculation, as the Ba and ^{226}Ra contents of the suspended sediment of the various rivers are not known. If Danube sediments are similar to those in other rivers, reducing the desorbed inputs of the dammed rivers alone is not sufficient to explain the Ba surface water deficit. Radium-226 has a higher ratio of dissolved to desorbed inputs than Ba (2:5 compared with 0.3:2)^{3–6,28}, and we would expect the decrease in sediment flux to be more

important for ^{226}Ra than Ba. Although this is in accordance with the greater depletion of ^{226}Ra from Black Sea surface waters, it must be noted that the observed ^{226}Ra decrease also covers a longer time.

We constructed a simple two-box model of the Black Sea to examine the relative importance of barite removal and riverine inputs to surface Ba and ^{226}Ra concentrations (Fig. 2). The salinity distribution together with the water budget²⁵ was used to constrain mixing between the surface (18 parts per thousand p.p.t.) and deep (22 p.p.t.) boxes. A range of mixing fluxes, obtained for the range of salinities (30–38 p.p.t.) assigned to Mediterranean waters flowing in at the Bosphorus, resulted in deep-water replacement times of 500–1,000 years, in accordance with other recent estimates²⁹. As deep-water Ba concentrations

FIG. 2 Schematic of box model with areas of arrows proportional to indicated water exchange rates²⁵. The boundary between surface and deep box is taken to be the pycnocline as this is where abrupt concentration changes in concentration are observed for detailed Ba profiles². Areas and volumes have been derived from bathymetric charts of the Black Sea. The surface area of the Black Sea excluding Sea of Azov²⁵ is $395,000 \text{ km}^2$. Surface and deep box exchange rates are given for a 38-p.p.t. Bosphorus input which has measured Ba and Ra-226 concentrations of 125 nM (ref. 2) and 16 d.p.m. per 100 l (ref. 1). The net inflow at the Kerch Straits was presumed to be riverine, and the exchange was assumed to be conservative. Deep-water Ba and Ra-226 concentrations are 460 nM (ref. 2) and 15 d.p.m. per 100 l (ref. 1). The particulate flux of Ba and Ra-226 is represented by F_d , and f is the fraction of that flux which returns to solution. The model does not specify whether dissolution occurs in deep box or in sediments². See text for further details.



have remained constant and mixing of surface waters is rapid, the steady-state assumption was made to obtain values for the first-order removal constants, both for the 1975 and 1988–1989 Ba distributions.

We estimated the riverine Ba endmember (300–450 nM) from measurements in the Danube, Dnepr, Dnestr, Kuban and Prut channels (K.K.F. and J.M.E., unpublished data), Danube Estuary³⁰ and in other rivers^{3,5,6}.

This exercise showed that the 1975 biological removal constant of $\sim 700 \text{ km}^3 \text{ yr}^{-1}$ must be roughly doubled to produce the surface Ba concentrations observed today. If the 1988–89 biological removal constant is applied to the 1975 distribution, an unreasonably high effective riverine concentration ($\geq 1,000 \text{ nM}$) is required, and using the 1975 removal term for the 1988–89 Ba situation would require the impossible situation that the rivers become a net sink for Ba. Thus the model indicates that biological forcing is primarily responsible for declining surface Ba concentrations in the Black Sea. Although the predicted flux of Ba from the surface box ($\sim 0.3 \text{ mg Ba m}^{-2} \text{ d}^{-1}$) is ten times the average value for the entire sediment trap deployment ($0.035 \text{ mg Ba m}^{-2} \text{ d}^{-1}$) (B. Hay and S. Honjo, personal communication), productivity is known to vary from zone to zone in the Black Sea, and the traps were located outside the most productive northern regions²⁶. Furthermore, to maintain higher concentrations of Ba in the deeper box, a large fraction ($\sim 70\%$) of barite must return to solution. The occurrence of dissolution in spite of 2.5 times supersaturation with respect to pure barite^{2,23} suggests that the trapped barite is a poorly crystalline and/or impure phase, possibly bearing potassium or calcium (T. Church, unpublished data).

Assuming that ^{226}Ra is incorporated into barite with a K_d of ~ 1 , as shown for the open ocean²², taking account of ^{226}Ra decay in the deep box and using the removal and dissolution terms determined for Ba, we calculate a riverine ^{226}Ra endmember of 16–18 d.p.m. per 100 l for the 1988–89 period. This falls in the range of values found for world rivers, but has yet to be directly characterized in Black Sea estuaries. A surface concentration of 21 d.p.m. ^{226}Ra per 100 l can be achieved by both lowering the biological removal constant by a factor of 5 and increasing the riverine contribution by 20%. If the mechanism linking barite and organic carbon fluxes can be clarified and riverine inputs determined, continued monitoring of surface water concentrations of Ba and ^{226}Ra could provide an integrated measure of the changing and otherwise difficult-to-quantify in net primary productivity in the Black Sea. □

Received 17 September 1990; accepted 26 February 1991.

- O'Neill, D. J., Moore, W. S. & Todd, J. F. *Earth Planet. Sci. Lett.* (in the press).
- Falkner, K. K. et al. *Earth planet. Sci. Lett.* (in the press).
- Honor, J. S. & Chan, L. H. *Earth planet. Sci. Lett.* **37**, 242–250 (1977).
- Li, Y.-H., Mathieu, G., Biscaye, P. & Simpson, H. *Earth planet. Sci. Lett.* **37**, 237–241 (1977).
- Edmond, J. M., Boyle, E. D., Drummond, D., Grant, B. & Mislick, T. *Neth. J. Sea Res.* **12**, 324–328 (1978).
- Li, Y.-H. & Chan, L. H. *Earth planet. Sci. Lett.* **43**, 343–350 (1979).
- Elsinger, R. J. & Moore, W. S. *Earth planet. Sci. Lett.* **48**, 239–249 (1980).
- Key, R. M., Stallard, R. S., Moore, W. S. & Sarmiento, J. L. *J. geophys. Res.* **90**, 6995–7004 (1985).
- Chan, L. H. et al. *Earth planet. Sci. Lett.* **32**, 258–267 (1976).
- Grashchenko, S. M., Nikolayev, D. S., Kolyadin, L. B., Kuznetsov, Y. V. & Lazarev, K. F. *Dokl. Akad. Nauk SSSR* **132**, 1171–1172 (1960).
- Moore, W. S. *Earth planet. Sci. Lett.* **6**, 437–446 (1969).
- O'Neill, D. J. thesis, Univ. South Carolina (1986).
- Klinkhammer, G. P. & Chan, L. H. *Analytica chim. Acta* **232**, 323–329 (1990).
- Butts, J., Todd, J. F., Lerch, I., Moore, W. S. & Moore, D. G. *Mar. Chem.* **25**, 349–357 (1988).
- Tolmazin, D. *Prog. Oceanogr.* **15**, 217–276 (1985).
- Boguslavskiy, S. G., Sarkisyan, A. S., Dzhiyev, T. Z. & Koveshnikov, L. A. *Atmos. ocean. Phys.* **12**, 337–340 (1976).
- Kaufman, A., Trier, R. M., Broecker, W. S. & Freely, H. W. *J. geophys. Res.* **78**, 8827–8848 (1973).
- Li, Y.-H., Freely, H. W. & Toggweiler, J. R. *Deep Sea Res.* **27A**, 545–555 (1980).
- Dehairs, F., Chesselet, R. & Jedwab, J. *Earth planet. Sci. Lett.* **49**, 528–550 (1980).
- Collier, R. & Edmond, J. *Prog. Oceanogr.* **13**, 113–119 (1984).
- Dehairs, F., Lambert, C. E., Chesselet, R. & Risler, N. *Biogeochemistry* **4**, 119–139 (1987).
- Moore, W. S. & Dymond, J. *Earth planet. Sci. Lett.* (in the press).
- Church, T. M. & Wolgemuth, K. *Earth planet. Sci. Lett.* **15**, 35–44 (1972).
- Bishop, J. K. B. *Nature* **332**, 341–343 (1988).
- Sorokin, Y. I. in *Ecosystems of the World* 26 (ed. Ketchum, B. H.) 253–292 (Elsevier, New York, 1983).
- Hay, B. J. et al. *Deep Sea Res.* **37**, 911–928 (1990).
- Denisova, A. I. *Vodnyye Resursy* **6**, 167–172 (1983).

- Elsinger, R. J. & Moore, W. S. *Estuar. Coast Shelf Sci.* **18**, 601–613 (1984).
- Top, Z., Izdar, E., Ergun, M. & Konuk, T. *EOS* **71**, 1020–1021 (1990).
- Klinkhammer, G. P. & Falkner, K. K. *EOS* **70**, 1150 (1989).

ACKNOWLEDGEMENTS. We thank our colleagues at the Institute of Biology of the Southern Seas at Sevastopol, I. Salihoglu, K. Beusseler, H. Livingston, S. Honjo, B. Hay, S. Manganini, G. Klinkhammer, K. Lange and S. Recca for their contributions to this study, and F. Dehairs and T. Church for their comments on the manuscript. This work was supported in part by the NSF and by a Vetlesen Grant from the Woods Hole Oceanographic Institution.

Melt droplet formation in energetic impact events

H. J. Melosh & A. M. Vickery

Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

IT has long been known that impact between rocky bodies at velocities of $\geq 15 \text{ km s}^{-1}$ can melt or vaporize both the impacting object and a portion of the target¹. We have recently shown^{2,3} that geological materials initially shocked to high pressure approach the liquid–vapour phase boundary from the liquid side as they decompress, breaking up into an expanding spray of liquid droplets. Here we present a simple theory for estimating the sizes of these droplets as a function of impactor size and velocity, and show that these sizes are consistent with observations of microtektites and spherules found in the Cretaceous/Tertiary boundary layer.

The size of melt droplets formed in an impact is important for the study of tektites and microtektites. Although there are still proponents of the theory that tektites are not the products of impacts⁴, the recent discovery of microtektite-size spherules at the Cretaceous/Tertiary boundary^{5,6} and in the ejecta from the Acraman impact structure⁷ clearly tie spherule formation to terrestrial impacts. The pronounced iridium anomaly associated with the spherule beds found in Archaean greenstone belts in South Africa^{8,9} also implies that impacts are capable of generating large volumes of microtektite-size spherules. Even at Meteor Crater, Arizona the bulk of the iron projectile seems to have been dispersed into a vast number of nickel–iron spherules¹⁰.

It has therefore become clear that spherule formation is an important part of the impact-cratering process, but until now no reliable method has been found for estimating the size of spherules in a given impact event. Previous studies that considered the break-up of melt streams or sheets either calculated droplet sizes from the break-up of thin cylindrical 'jets' of melt¹¹ or incorrectly assumed that the minimum droplet size is inversely proportional to the ejection velocity (rather than to the differential velocity within the melt), and calculated the maximum droplet size for melt interacting with gas (either ambient atmosphere or impact-generated vapour) for an arbitrary range of differential velocities between the two phases¹². Other previous work considered the homogeneous condensation of a vapour phase^{12,13}. Our recent studies of the decompression of shocked dunite^{2,3,14} and quartz using the ANEOS equation-of-state program¹⁵ show that at common asteroidal impact velocities on Earth ($\sim 20 \text{ km s}^{-1}$), the shocked material follows a release path that intersects the liquid–vapour phase curve from the liquid side (see Fig. 1). Droplet formation by homogeneous condensation from a vapour is thus not an important feature of asteroidal impacts on the Earth, although it may be important for cometary impacts. Here we present a simple theory of melt droplet formation from the liquid phase in impacts. The theory is not intended to be rigorous, as detailed data for testing a rigorous theory do not exist, but is presented as an order-of-magnitude theory which, we hope, captures the essential physical processes occurring during droplet formation.

Shortly after an energetic impact event, the impacting projectile and a roughly equal volume of the target are compressed to high pressure¹⁶. For most solids (including rocks) the density

of the strongly shocked material is about twice the zero-pressure density. On reaching a free surface, the compressive shock waves are reflected as rarefaction waves which decompress the previously shocked material to low pressure. The net result of shock compression and release is a thermodynamic cycle that does work on the material and ejects it from the growing crater. In general, projectile material is ejected first and attains the highest speed, whereas target material is ejected later and is slower. The shock level of target material ranges from the highest achieved in the impact down to nearly zero.

For the impact of silicate projectiles on Earth, the most highly shocked material begins in the liquid phase and expands to lower pressure and density. At this time, the liquid undergoes an initial fragmentation into clumps whose diameter d_0 is determined¹⁷ by the balance between surface tension σ and the local kinetic energy, $1/2mv^2$, where $v = \dot{\epsilon}d_0$. This balance yields an initial clump size of

$$d_0 = \left(\frac{40\sigma}{\rho_l \dot{\epsilon}^2} \right)^{1/3} \quad (1)$$

where ρ_l is the liquid density at the moment of fragmentation and $\dot{\epsilon}$ is the strain rate of the expanding melt. (Note that because the clump size depends on the local strain rate, the geometry of melt ejection is irrelevant as long as the thickness of the melt sheet or stream is large compared to the inferred d_0 .) For silicate liquids, σ is typically $\sim 0.3 \text{ N m}^{-1}$ (although σ decreases with increasing temperature and vanishes at the critical point¹⁸), ρ_l is about the uncompressed density, $2,500 \text{ kg m}^{-3}$, and $\dot{\epsilon}$ can be estimated from the mean velocity of expansion of the melt plume, v_{exp} , and its diameter R_f at the time of melt fragmentation, $\dot{\epsilon} \approx v_{\text{exp}}/R_f$. We take $R_f \approx L/2$, where L is the projectile diameter and estimate v_{exp} from enthalpy conservation in the expanding plume as $v_{\text{exp}} = \sqrt{2E}$, where $E \approx 1/2v_p^2 \approx 1/8v_i^2$; v_p is the particle velocity in the target, and v_i is the impact velocity. Thus, $v_{\text{exp}} \approx v_i/2$, that is, the mean velocity of the expanding melt plume is about half of the impact velocity itself. This is in fact an overestimate of the velocity, as the melt plume at this early stage will not yet have achieved its final velocity, so that the actual clump size will be larger by some fraction than computed. Fortunately the cube root in equation (1) makes the result rather insensitive to such quibbles, and we achieve as a final estimate $d_0 \approx (40\sigma/\rho_l)^{1/3}(L/v_i)^{2/3}$. For a 1-km-diameter projectile with an impact velocity $v_i = 20 \text{ km s}^{-1}$, $d_0 \approx 2 \text{ cm}$, which is comfortably close to the size of tektites, supposing that some of the melt ejected from the crater escapes further fragmentation. This is more likely to occur in the lower speed, inner portion of the ejecta (which is not included in our model) comprising mostly target material, consistent with the lack of siderophile element (projectile) contamination of tektites⁴.

Following the initial fragmentation of the homogeneous melt, a second generation of droplet formation occurs in the more energetic part of the ejecta plume, when the material reaches the liquid-vapour phase boundary and begins to vaporize. The ultimate droplet size is thus established in an environment in which both vapour and melt droplets are present. At this time the ejecta plume is still accelerating under the influence of pressure gradients established during earlier phases of the impact, and it is expected that the gas will be moving outward faster than the melt droplets. Indeed, once melt droplets have formed they can only be accelerated by the drag between themselves and the gas.

When the volume of vapour is much greater than the volume of melt, the droplet radius r is estimated from the balance of aerodynamic forces and surface tension. Break-up occurs at a critical Weber number¹⁹ of ~ 10 , hence

$$r = \frac{5\sigma}{\rho_g(v_g - v_d)^2} \quad (2)$$

where ρ_g is the vapour density and v_g and v_d are the velocities of the vapour and droplet, respectively. As the gas streams past the droplets it imparts an acceleration due to gas drag

$$\frac{dv_d}{dt} = \frac{3C_D}{4} \frac{\rho_g}{\rho_d} \frac{(v_g - v_d)^2}{r} = \frac{3C_D}{20\sigma} \rho_g^2 (v_g - v_d)^4 \quad (3)$$

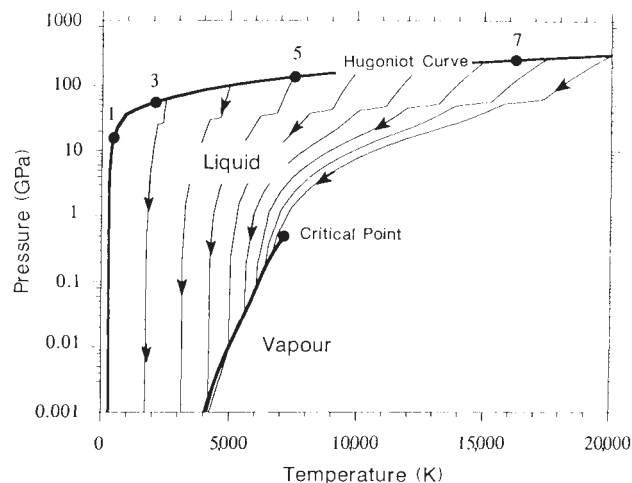
where C_D is the drag coefficient (assumed to be constant, which implies that velocities are high enough to be in the turbulent flow regime) and ρ_d is the droplet density. In the second step of (3) we presume that the drops break up until an equilibrium size is reached. This equilibrium is achieved when the acceleration of the droplets equals the acceleration of the gas. At this point, the gas and the droplets have a constant differential velocity $\Delta v = v_g - v_d$. Assuming that the mass fraction of liquid l remains constant during the break-up process, momentum balance requires

$$(1-l) \frac{dv_g}{dt} + l \frac{dv_d}{dt} = -\frac{1}{\rho} \frac{dP}{dr} = a \quad (4)$$

where ρ is the overall density of vapour plus liquid, a is the overall acceleration and dP/dr is the pressure gradient, which is assumed to be a constant given by the conditions of the hydrodynamic expansion. Although the theory can be amended to take account of rapid phase changes, we show that this is generally unnecessary.

Equations (3) and (4) are a coupled pair of first-order non-linear differential equations which can be solved analytically subject to the initial conditions that $v_g = v_d = 0$. They are most easily solved in terms of the differential velocity Δv and the

FIG. 1. The Hugoniot (heavy curve) and release adiabats (solid curves with arrows) in pressure-temperature space for quartz (SiO_2) computed by the ANEOS equation-of-state package. The liquid-vapour phase curve (heavy line) and the critical point are also shown. The equation of state includes the quartz-stishovite phase transformation at 14 GPa and yields a good fit to the Hugoniot data up to 650 GPa. The numbers above the Hugoniot curve give the particle velocity at the adjacent dots in km s^{-1} .



mean velocity $(v_g + v_d)/2$. The ultimate droplet radius (as $t \rightarrow \infty$) is

$$r_\infty = \left(\frac{15}{4} \frac{\sigma C_D}{\rho_d a} \right)^{1/2} \quad (5)$$

and the equilibrium differential velocity is

$$\Delta v = \sqrt{\frac{5\sigma}{\rho_g r_\infty}} \quad (6)$$

The solution reaches these asymptotic values on a timescale

$$t_{eq} = \frac{(1-l)\Delta v}{a} \quad (7)$$

Note that the equilibrium droplet size and differential velocity do not depend on the liquid fraction l so long as it remains approximately constant during the time required to achieve equilibrium. In the break-up process the droplet size distribution will be strongly peaked around the mean, as any droplets much larger than average lag behind and move more rapidly with respect to the gas and so break up, whereas smaller droplets move faster than average and soon run into other droplets. The droplet size will not change greatly as the gas expands and condenses on the existing drops, as the vapour fraction never exceeds ~ 0.5 , corresponding to a fractional droplet diameter increase of $(2)^{1/3} = 1.26$. Even if the impact velocity were high enough for the vapour fraction to be 0.9, the final droplet sizes would only be at most about a factor of two larger.

Equations (5) to (7) can be used on a cell-by-cell basis in detailed hydrocode computations to determine the droplet size distribution for a given impact event. But for most purposes only an approximate overall droplet size estimate is appropriate, so we may make further simplifications. To estimate r_∞ we need the acceleration at the time of droplet formation, $a = -(dP/dr)/\rho$. In an impact the acceleration declines monotonically with time, so that the smallest droplet size is determined at the first moment the vapour concentration becomes appreciable. At later times the acceleration is more gentle, and although larger droplets could exist in the slower vapour flow, we assume that the overall expansion prevents reaggregation. Approximating the plume as an expanding hemisphere of radius $R(t)$, our hydrocode results³ suggest that the vapour curve is first reached when the plume density ρ is one-fourth the maximum density, ρ_{max} , or one-half the uncompressed density ρ_0 , hence the plume radius R is $(4)^{1/3} R_0$. The Zel'dovich and Razier model for gas expansion²⁰ gives a parabolic distribution of pressure within the plume, $P(r) = P_c[1 - (r/R)^2]$, where $P_c = P_{max}(R_0/R)^{3\gamma}$. P_c is the pressure at $r = 0$ and P_{max} is the initial (maximum) pressure. The effective ratio of specific heats computed from ANEOS is $\gamma = 1.8$. By the Hugoniot relations, $P_{max} = \rho_0 U_s u_p$, where U_s is the shock velocity and u_p is the particle velocity. The relation between U_s and u_p is given approximately by $U_s = c + s u_p$, where c and s are material constants. For projectile and target of identical materials $u_{p,max} \approx v_i/2$. Assuming that R_0 is one-half of the impactor diameter L , we find that the acceleration at the time of droplet formation in the outer portion of the plume is given by $a \approx 0.1 v_i(2c + s v_i)/L$. For quartz, $c = 1.74 \text{ km s}^{-1}$ and $s = 1.7$ (ref. 21), so for asteroidal impact velocities, $s v_i \gg 2c$ and $a \approx 0.1 s v_i^2/L$. Substituting this into equation (5) and using $\sigma = 0.3 \text{ N m}^{-1}$, $C_D = 1$, and $\rho_d \approx 10^3 \text{ kg m}^{-3}$, the equilibrium droplet size is given by

$$r_\infty \approx 0.11 \frac{\sqrt{L}}{v_i} \quad (\text{SI units}) \quad (8)$$

which should be valid for the high-energy portion of the ejecta for impact velocities not too different from 20 km s^{-1} . (For lower energy portions of the ejecta, or for impact velocities low enough so that there is little or no vaporization during release, equation (1) applies. For impact velocities high enough that the entropy of the most highly shocked ejecta exceeds the entropy of the

critical point ($v_i > 30 \text{ km s}^{-1}$), homogeneous nucleation from a vapour may be important.) This is a minimum size insofar as droplets formed in the slower-accelerating cloud interior or at a later stage of expansion will be substantially larger.

The impact that deposited the K/T boundary layer is believed to have involved a projectile of diameter $\sim 10 \text{ km}$. For $v_i = 20 \text{ km s}^{-1}$, we estimate $r_\infty = 550 \mu\text{m}$, consistent with the observed spherule size in the K/T boundary layer at Caravaca, Spain and Stevns Klint, Denmark⁵. Equations (6) and (7) predict a differential velocity $\Delta v \approx 1.3 \text{ m s}^{-1}$ and a droplet formation time of 0.05 ms , far smaller than the expansion time of $\sim 800 \text{ ms}$. It can be shown, in general, that for all impacts of geological interest the droplet formation time (7) is much less than the expansion time R/v_i , so that equations (5)–(8) are generally valid. Furthermore, the differential velocity is never large enough for macroscopic separation of droplets and gas to be important.

The smallest reported size fraction of the late Eocene microtektites²² is $63\text{--}125 \mu\text{m}$; inverting equation (8), we estimate a projectile diameter L of $\sim 130\text{--}330 \text{ m}$ for a 20 km s^{-1} impact velocity. This late Eocene event was much smaller in magnitude than the K/T event as indicated by the much more restricted size of the strewn field and the fact that there are no extinctions associated with it^{23,24}, so this much smaller impactor size seems reasonable.

In some cases, the minimum reported spherule sizes are inconsistent with the predictions of our model for reasonable values of L and v_i . The microspherule beds of the Archaean greenstone belt, for instance, contain spherules up to 4 mm in size and contain a truly impressive total mass of spherules: bed S2, for instance, is from 1 to 3 m thick and extends in outcrop over an area of at least $20 \times 50 \text{ km}^2$ (ref. 9). Although these beds are probably reworked, and the original thickness may have been only several decimetres (J. Smit, personal communication), this mass of spherules implies an impactor much larger than the K/T projectile. Using the maximum spherule size, our model implies a projectile of $\sim 160 \text{ km}$ in diameter, 16 times larger than the K/T projectile. The minimum spherule size, however, is only $100 \mu\text{m}$, similar to the late Eocene microtektites.

There are a number of reasons why impact-produced spherules may be smaller than predicted by our formulation. First, the surface tension decreases with increasing temperature, going to zero at the critical point, and a smaller value for σ reduces the predicted droplet size. Second, the melt droplets may have been subjected to a further stage of break-up when the expanding melt and vapour plume from the impact came into contact with the ambient air, establishing a strong decelerating dP/dr gradient. Third, friction with the atmosphere after the droplets decouple from the plume may cause ablation, and the ablated portions would form much smaller drops than the 'parent' spherule. Detailed numerical computations of the interaction of the projectile plume with the surrounding atmosphere should be able to answer these questions. $\square$

Received 9 November 1990; accepted 28 February 1991.

- Ahrens, T. J. & O'Keefe, J. D. *The Moon* **4**, 214–249 (1972).
- Melosh, H. J. in *Origin of the Earth* (eds Jones, J. H. & Newsom, H. E.) 69–83 (Oxford University Press, New York, 1990).
- Vickery, A. M. *Eos* **71**, 1429 (1990).
- Koeberl, C. A. *Rev. Earth planet. Sci.* **14**, 323–350 (1986).
- Smit, J. & Klaver, G. *Nature* **292**, 47–49 (1981).
- Montanari, A. et al. *Geology* **11**, 668–671 (1983).
- Wallace, M. W., Gostin, V. A. & Keays, R. R. *Meteoritics* **25**, 161–165 (1990).
- Lowe, D. R. & Byerly, G. R. *Geology* **14**, 83–86 (1986).
- Lowe, D. R., Byerly, G. R., Asaro, F. & Kyte, F. J. *Science* **245**, 959–962 (1989).
- Nininger, H. H. *Arizona's Meteorite Crater* 1–232 (American Meteorite Laboratory, Denver, 1956).
- Scarlett, B. & Buxton, R. E. *Earth planet. Sci. Lett.* **22**, 177–187 (1974).
- O'Keefe, J. D. & Ahrens, T. J. in *Spec. Pap. geol. Soc. Am.* **190**, 103–120 (1982).
- Melosh, H. J. in *Spec. Pap. geol. Soc. Am.* **190**, 121–127 (1982).
- Benz, W., Cameron, A. G. W. & Melosh, H. J. *Icarus* **81**, 113–131 (1989).
- Thompson, S. L. & Lauson, H. S. *Improvements in the Chart D Radiation-hydrodynamic Code III: Revised Analytic Equations of State* (Sandia National Laboratory, Albuquerque, 1972).
- Melosh, H. J. *Impact Cratering: A Geologic Process* 1–245 (Oxford University Press, 1989).
- Grady, D. E. *J. appl. Phys.* **53**, 322–325 (1982).
- Brunauer, S. & Copeland, L. E. in *Handbook of Physics* (eds Condon, E. U. & Odishaw, H.) 5.90–5.115 (McGraw-Hill, New York, 1967).

19. Young, G. A. *The Physics of the Base Surge* (US Naval Ordnance Laboratory, White Oak, 1965).
20. Zel'dovich, Y. B. & Raizer, Y. P. *The Physics of Shock waves and High Temperature Hydrodynamic Phenomena* (Academic, New York, 1967).
21. Wackerle, J. J. *appl. Phys.* **33**, 922-937 (1962).
22. Glass, B. P., DuBois, D. L. & Ganapathy, R. *J. geophys. Res.* **90**, 425-428 (1982).
23. Glass, B. P., Burns, C. A., Crosbie, J. R. & DuBois, D. L. *J. geophys. Res.* **90**, 175-196 (1985).
24. Sanfilippo, A., Riedel, W. R., Glass, B. P. & Kyte, F. T. *Nature* **314**, 613-615 (1985).

The viceroy butterfly is not a batesian mimic

David B. Rittland & Lincoln P. Brower

Department of Zoology, University of Florida, Gainesville, Florida 32611, USA

DEFENSIVE mimicry has long been a paradigm of adaptive evolution by natural selection¹⁻³. Mimics, models and predators in a batesian mimicry system (unpalatable model, palatable mimic) exist in a very different selective milieu from those in a müllerian system (involving ≥ 2 unpalatable 'co-models')^{1,4-6}. Consequently, the incorrect characterization of a mimicry relationship obscures the natural histories of populations involved and undermines attempts to test general mimicry theory by means of empirical studies of specific systems. Here, we reassess the classic case of mimicry involving viceroy butterflies, *Limenitis archippus* (Cramer) (Nymphalidae), and two species they purportedly mimic: the monarch, *Danaus plexippus* (L.), and the queen, *Danaus gilippus* (Cramer) (Nymphalidae: Danainae). Viceroy is historically considered palatable (batesian) mimics^{7,8} of the chemically defended⁹ danaines. Our experiment refutes this interpretation by revealing that viceroys are as unpalatable as monarchs, and significantly more unpalatable than queens from representative Florida populations. This implies that viceroys are müllerian co-mimics of the danaines and prompts a comprehensive reassessment of this widely cited exemplar of mimicry.

The established conception of viceroy-danaine mimicry as batesian has been challenged by (1) suggestions that monarchs and queens sometimes possess insufficient chemical defences (cardenolides^{10,11} and pyrrolizidine alkaloids^{12,13}) to serve as models, and (2) indirect evidence¹⁴⁻¹⁶ suggesting that viceroys are themselves weakly unpalatable. To clarify the status of viceroy-danaine mimicry, we directly quantified the palatability of Florida viceroys in relation to both monarchs and queens. We hypothesized that viceroys may be involved in two quite different mimicry relationships with these danaines, because monarchs and queens differ in their capacities to store cardenolides¹⁰ and perhaps in their pyrrolizidine alkaloid content^{13,17}.

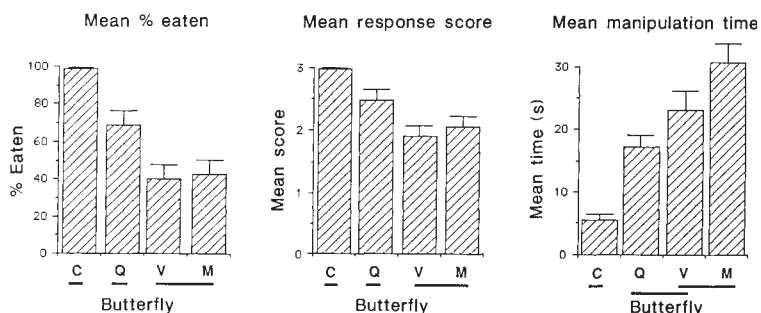
Our study employed a rigorous bioassay that separates unpalatability *per se* from preconditioned visual or gustatory aversions, and alleviates several additional experimental prob-

lems: predators' previous experience, aversions learned during the experiment, neophobia and the effects of presentation order. Wild-caught male red-winged blackbirds (*Agelaius phoeniceus* (L.); Icteridae) were used as ecologically relevant predators: they are abundant year-round residents in Florida, include substantial animal matter in their diet, forage in marshy areas where viceroys and danaines occur¹⁸ and attack butterflies in the wild¹⁹. We sampled butterflies that, on the basis of available larval and adult foodplants, are assumed to be representative of many Florida populations in terms of chemical defence and palatability²⁰⁻²². We presented only abdomens to the birds, thus compelling them to accept or reject butterflies based primarily on gustatory cues, rather than on wing pattern cues that they might have encountered in the wild. (Additional experiments (unpublished results) demonstrate that in this system, the relative palatability of abdomens reasonably estimates that of entire butterflies.) Palatabilities of viceroy, queen and monarch abdomens were compared to one another and to four palatable control species (see legend to Fig. 1). Every bird received each type of butterfly abdomen and therefore served as its own control. When an abdomen was presented, we assessed the bird's reaction by recording a 'response score' (see legend to Fig. 1) and, if the bird ate the entire abdomen, the manipulation time (time elapsed between attack and swallowing). Throughout the experiment control abdomens were treated similarly by the birds (no significant difference in the percentage eaten, response score or manipulation time), and therefore control abdomens are considered as a single group in the following treatment.

Palatability indices for control and experimental abdomens, summarized in Fig. 1 (pooled) and Table 1 (by bird), indicate that viceroys were significantly less palatable than the controls. Only 41% of viceroys were totally eaten (compared with 98% of controls), and 35% were taste-rejected after a single peck, yielding significantly lower response scores. Only two birds (12.5%) ate six or more of the eight viceroys offered. Birds commonly exhibited distress behaviour (head-shaking, excessive drinking, agitation) after tasting a viceroy abdomen, and abdomens that were eventually eaten were generally accepted very hesitantly by the birds, with a mean manipulation time nearly five times that of controls. These data clearly refute the traditional hypothesis that viceroys are palatable batesian mimics.

Moreover, although both the monarch and queen were moderately unpalatable (Fig. 1, and Table 1), neither exceeded the unpalatability of its purported mimic, the viceroy. In fact, queens were significantly more palatable than viceroys (68% eaten and only 10% taste-rejected), whereas monarchs and viceroys were about equally unpalatable (Fig. 1). We therefore conclude that both the viceroy/monarch and viceroy/queen relationships exemplify müllerian mimicry, and that viceroys are co-models of the danaines rather than parasitic mimics. But the two relationships differ. The similar unpalatability of viceroys and monarchs suggests that their relationship (in sampled populations) exemplifies classic, symmetrical müllerian mimicry. By

FIG. 1 Mean palatability indices (with standard error) derived from responses of 16 red-winged blackbirds to abdomens of control (C), queen (Q), viceroy (V) and monarch (M) butterflies. Means joined by a line do not differ significantly (α of 0.05, Friedman two-way analysis and Joint-rank Ryan multiple comparison²⁷). On each of six consecutive days, each bird received four control and four experimental (viceroy or danaine) abdomens in random order (individually presented for 12-s at 10-min intervals). Overall, each bird received 8 viceroy, 8 queen and 8 monarch abdomens (in different, randomly determined orders) interspersed with 24 control abdomens. Mean response score was calculated from scores recorded for individual presentations: 0, not touched (sight-rejected); 1, pecked once (taste-rejected); 1.5, pecked repeatedly, but not eaten; 2, partly eaten; 3, entirely eaten. Butterfly sources and probable larval foodplants: viceroys, Dixie/Alachua Co., Florida (*Salix caroliniana* Michx.); queens, Collier/Hendry Co., Florida (*Sarcostemma clausum* R. & S., *Asclepias*



incarnata L.; monarchs (autumn migrants), Dixie/Alachua Co., Florida (*Asclepias syriaca* L.). Control butterfly species: *Anartia jatrophae* (Johannson), *Phoebis sennae* (L.), *Papilio glaucus* L., *Papilio palamedes* Drury.

TABLE 1 Responses of birds to butterflies

Bird	Number entirely eaten				Percentage entirely eaten				Mean response score				Mean manipulation time			
	C	Q	V	M	C	Q	V	M	C	Q	V	M	C	Q	V	M
A	22	0	1	0	92	0	12	0	2.92	0.50	0.88	1.25	5.2	—	—	—
B	24	6	5	5	100	75	62	62	3.00	2.62	2.25	2.50	5.3	18.9	12.4	25.9
C	24	5	5	4	100	62	62	50	3.00	2.44	2.50	2.12	8.2	25.2	26.6	19.6
D	24	7	4	5	100	88	50	62	3.00	2.88	2.12	2.25	3.1	12.4	41.6	27.4
E	24	5	0	4	100	62	0	50	3.00	2.69	1.06	2.38	8.6	36.5	—	58.5
F	24	7	2	2	100	88	25	25	3.00	2.75	1.62	1.38	8.7	16.4	47.0	25.0
G	23	5	3	3	96	62	38	38	2.96	2.50	1.94	1.94	4.5	18.1	24.2	30.0
H	24	6	5	3	100	75	62	38	3.00	2.62	2.62	2.19	3.8	10.0	35.3	32.1
I	24	6	8	1	100	75	100	12	3.00	2.75	3.00	1.62	3.9	11.4	13.2	49.4
J	24	6	2	8	100	75	25	100	3.00	2.69	1.69	3.00	3.2	13.5	22.1	15.2
K	24	8	2	5	100	100	25	62	3.00	3.00	1.81	2.38	3.4	16.5	15.4	18.7
L	22	4	2	2	92	50	25	25	2.92	2.00	1.88	1.25	4.7	21.0	20.8	30.0
M	23	4	4	3	96	50	50	38	2.92	2.50	2.50	2.38	9.1	25.4	23.9	39.1
N	24	8	4	8	100	100	50	100	3.00	3.00	2.00	3.00	4.6	7.6	7.5	23.7
O	24	4	6	2	100	50	75	25	3.00	2.50	2.38	2.00	5.2	15.7	15.0	42.6
P	23	6	0	4	96	75	0	50	2.92	2.50	1.38	2.00	7.0	13.6	—	32.6
$\bar{x}$	23.6	5.4	3.3	3.7	98	68	41	46	2.98	2.50	1.98	2.10	5.3	17.5	23.5	31.3
s.e.	0.2	0.5	0.6	0.6	1	6	7	7	0.02	0.15	0.14	0.13	0.5	1.9	3.3	3.0

Individual responses of 16 red-winged blackbirds to control (C), queen (Q), viceroy (V) and monarch (M) abdomens. Means are compared statistically in Fig. 1. Response score described in Fig. 1; manipulation time is in seconds (birds eating no abdomens of a given species have no manipulation time recorded). Individual birds differed in their treatment of viceroys and danaines; this is an important aspect of mimicry dynamics.

contrast, the viceroy/queen relationship is asymmetrical, with the viceroy a stronger (more unpalatable) co-model than the queen. Mathematical models of müllerian mimicry suggest that the queen actually benefits more from the mimicry than does the viceroy²³, perhaps even at the viceroy's expense²⁴, reversing the traditionally accepted roles of model and mimic.

Our laboratory study shows reactions of one predator species to abdomens from representative Florida populations; extrapolating to natural interactions involving multiple predators and variably palatable butterflies must obviously be done with caution. Nonetheless, our finding of müllerian mimicry prompts a reinterpretation of viceroy-danaine-predator interactions. The classic Batesian interpretation of viceroy-danaine mimicry predicts that only the viceroy benefits, at the expense of both danaines (which it 'parasitizes') and predator (which it deceives into eschewing palatable prey). By contrast, the new müllerian interpretation predicts that all three species benefit from enhanced predator learning, which occurs because of the viceroy and danaine co-models' shared aposematic signal. Our findings therefore imply that the butterflies and their predators interact in a selective milieu very different from that previously assumed, and suggest that other 'classic' mimicry relationships warrant closer scrutiny.

Further investigation of viceroy mimicry will not only clarify the nature of this particular system, but will also potentially give insights into the dynamic processes of selective predation that shape mimicry relationships in general. The viceroy-danaine system offers an opportunity to test mathematical and conceptual models^{6,23} of dynamic müllerian mimicry, and provides an interesting comparison to previous studies of müllerian mimicry in which the co-mimics are more closely related (for example, *Heliconius* butterflies²⁵ or *Zygaena* moths²⁶). Thus, over a century after its discovery, the viceroy-danaine relationship continues to contribute to our understanding of the evolution and ecology of mimicry. □

Received 1 October 1990; accepted 11 January 1991.

1. Turner, J. R. G. *Ecol. Ent.* **12**, 81–95 (1987).
2. Vane-Wright, R. I. *Biol. J. Linn. Soc. Lond.* **16**, 33–40 (1981).
3. Brower, L. P. *Am. Nat.* **131** (suppl.), 1–3 (1988).
4. Ford, E. B. *Ecological Genetics* (Methuen, London, 1971).
5. Rothschild, M. *Symb. Bot. Uppsala* **22**, 82–99 (1979).
6. Huheey, J. E. *Am. Nat.* **131**, S22–S41 (1988).
7. Walsh, B. D. & Riley, C. V. *Am. Ent.* **1**, 189–193 (1869).
8. Scudder, S. H. *Nature* **3**, 147 (1970).
9. Brower, L. P. in *The Biology of Butterflies* (eds Vane-Wright, R. I. & Ackery, P. R.) 109–135 (Symp. Roy. Ent. Soc. No. 11).

10. Cohen, J. A. *J. chem. Ecol.* **11**, 85–103 (1985).
11. Malcolm, S. B., Cockrell, B. J. & Brower, L. P. *J. chem. Ecol.* **15**, 819–854 (1989).
12. Glendinning, J. I. thesis, Univ. of Florida (1989).
13. Kelley, R. B., Seiber, J. N., Jones, A. D., Segall, H. J. & Brower, L. P. *Experientia* **43**, 943–946 (1987).
14. Brower, J. V. Z. *Evolution* **12**, 32–47 (1958).
15. Brower, J. V. Z. *Evolution* **12**, 273–285 (1958).
16. Platt, A. P., Coppinger, R. P. & Brower, L. P. *Evolution* **25**, 692–701 (1971).
17. Pliske, T. E. *Environ. Ent.* **4**, 455–473 (1975).
18. Bent, A. C. *Life Histories of North American Blackbirds, Orioles, Tanagers, and Allies* (Dover, New York, 1965).
19. Brower, L. P. & Brower, J. V. Z. *Nat. Hist.* **71**, 8–19 (1962).
20. Rothschild, M., Mummery, R. & Farrell, C. *Biol. J. Linn. Soc. Lond.* **28**, 359–372 (1986).
21. Ackery, P. R. & Vane-Wright, R. I. *Milkweed Butterflies: Their Cladistics and Biology* (Cornell Univ. Press, New York, 1984).
22. Wunderlin, R. P. *Guide to the Vascular Plants of Central Florida* (University Presses of Florida, Tampa, 1982).
23. Turner, J. R. G., Kearney, E. P. & Exton, L. S. *Biol. J. Linn. Soc. Lond.* **23**, 247–268 (1984).
24. Huheey, J. E. *Evolution* **30**, 86–93 (1976).
25. Turner, J. R. G. *Zool. J. Linn. Soc.* **58**, 297–308 (1976).
26. Sbordoni, V., Bullini, L., Scarpelli, G., Forestiero, S. & Rampini, M. *Ecol. Ent.* **4**, 83–93 (1979).
27. Day, R. W. & Quinn, G. P. *Ecol. Monogr.* **59**, 433–463 (1989).

Left neglect for near but not far space in man

Peter W. Halligan & John C. Marshall

Neuropsychology Unit, University Department of Clinical Neurology, The Radcliffe Infirmary, Oxford OX2 6HE, and Rivermead Rehabilitation Centre, Oxford OX1 4XD, UK

IT has been suggested that, among the many visual areas of the human brain, there might be one set of spatial maps specialized for 'near' (peripersonal) and another for 'far' (extrapersonal) space. A distinction between 'grasping distance' and 'walking distance'¹, or between a 'reaching field' and a pointing or throwing field² has commonly been made. Evidence for such a division has been found in monkeys. Unilateral ablation of the frontal eye field (area 8) produces a more prominent inattention (or 'neglect') for objects in contralesional far space than in near space; by contrast, unilateral ablation of frontal area 6, which receives direct projections from area 7b (the rostral part of the inferior parietal lobules) results in inattention to visual stimuli limited to contralesional near space³. Despite predictions that comparable dissociations should be found in man⁴, there has been no convincing evidence. We report here such evidence in a patient with a unilateral right hemisphere stroke. Within peripersonal space, he showed severe left visuo-spatial neglect on conventional tests, including the highly

TABLE 1 Transection displacements in millimetres from true centre

Step		1 Near space (pen)	2 Far space (light pen)	3 Far space (darts)	4 Near space (pen)	5 Near space (light pen)
Line length (cm)						
Near	Far	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)
30.5	166	+46.3 (24.8)	-4.3 (3.8)	+22.5 (4.8)	+64.3 (22.7)	+38.3 (9.7)
25.4	138	+64.7 (18.9)	+12.5 (15.5)	+11.7 (4.7)	+43.2 (19.9)	+27.3 (19.3)
20.3	114	+37.0 (12.1)	-10.0 (4.2)	+19.0 (4.3)	+33.0 (6.5)	+29.9 (9.7)
15.2	83	+26.5 (6.7)	-2.3 (4.9)	+1.0 (6.5)	+27.7 (10.1)	+12.8 (7.5)
10.2	56	+18.7 (8.9)	-3.3 (3.4)	+0.8 (3.9)	+24.3 (5.5)	-0.3 (8.2)
5.1	27	+4.8 (2.7)	-5.4 (3.2)	-3.8 (4.5)	+2.8 (1.7)	-6.1 (6.9)

Line bisection performance in the five experimental conditions: step 1, displacement from true centre at viewing distance 450 mm (bisection with a pen); step 2, displacement at viewing distance 2.44 m (bisection with a light-pen); step 3, displacement at viewing distance 2.44 m (bisection by dart-throwing); step 4, repetition of step 1; step 5, displacement at viewing distance 450 mm (bisection with a light-pen, six weeks later). The best-fitting linear equations for transection displacement as a function of line length (expressed as visual angle) are: 1, $-3.46 \text{ mm} + (1.630 \times \text{visual angle})$ (81%); 2, $-6.57 \text{ mm} + (0.198 \times \text{visual angle})$ (9%); 3, $-10.25 \text{ mm} + (0.840 \times \text{visual angle})$ (83%); 4, $-5.09 \text{ mm} + (1.683 \times \text{visual angle})$ (92%); 5, $-16.15 \text{ mm} + (1.481 \times \text{visual angle})$ (95%). The final figures in brackets are the percentages of variance captured.

sensitive task of line bisection. When line bisection was performed in extrapersonal space, neglect was abolished or attenuated.

T.M. is a 57-year-old, right-handed mechanic who suffered a right hemisphere stroke on 1 July 1990. On admission (3 October 1990), T.M. had a severe left hemiparesis and a left inferior homonymous quadrantanopia; there was a full range of ocular movements and visual acuity was 6/9 in both eyes. Magnetic resonance imaging of the brain (29 November 1990) showed a large area of abnormality in the right hemisphere, with no signs of left hemisphere damage. Within the territory of the middle cerebral artery, his posterior parietal cortex is infarcted with some sparing of superior and medial parietal cortex (but the latter is undercut by deep white matter lesions). Most of the lateral and medial temporal cortex is affected, with dilation of the right temporal horn. There is also damage to the right cerebral peduncle, pons and internal capsule. The frontal pole and frontal cortex are spared, although the lateral inferior aspect of the right frontal lobe is affected. The occipital cortex is intact (hence the quadrantanopia is due to a tract lesion). Consistent with an intact left hemisphere, T.M. had a verbal intelligence quotient (109) that was within normal limits.

The outstanding neuropsychological symptom was left visuo-spatial neglect. Assessed on the Behavioural Inattention Test⁵, T.M. scored 71/146. This aggregate is derived from performance on: line- (23/36), letter- (17/40), and star-cancellation (29/54); line bisection (0/9); figure copying (0/4); and representational drawing to verbal request (2/3). The cut-off for normality is 130/146. On another line bisection task, where 11 horizontal lines between 25 mm and 279 mm are each presented 10 times in random order for bisection, T.M. performed very poorly. The mean of his transections was displaced rightwards at all line lengths; the linear regression of transection displacement on line length was $-6.23 \text{ mm} + (0.3194 \times \text{line length})$, and accounted for 99% of the variance. The multiplier (+0.3194) is over an order of magnitude larger than that in any member of a sample of 20 normal control subjects⁶. T.M. thus manifested classical left visuo-spatial neglect within peripersonal space. He did not show personal neglect. On verbal command, T.M. could touch left-sided parts of his body (elbow, shoulder, wrist, ear, thumb, knee, calf and thigh) with his right hand. Performance was without error, both with eyes open and eyes closed. The observations that follow were made during one testing session, 20 weeks after the stroke. Because of his hemiplegia, T.M. remained in his wheelchair throughout the session. Normal room illumination was maintained.

The investigations were conducted in the following order. Step 1: Black horizontal lines (2 mm thick) were individually drawn on sheets of paper (297 × 420 mm). Six line lengths were used; 51, 102, 152, 203, 254 and 305 mm. Each line, presented on a separate sheet, was centred on the page both horizontally

and vertically. The midpoint of each sheet was positioned at eye-level on a vertically oriented board directly in front of the patient. Viewing distance was ~450 mm, well within peripersonal space, and the visual angles subtended by the stimuli accordingly ranged between 6.4° (51-mm line) and 37.6° (305-mm line). The patient was instructed to mark the midpoint of each line with a fine pen held in the right hand. Step 2: A vertically oriented white board, 1.83 m long by 0.61 m high, was positioned so that the centre of the board was at eye-level for the patient in the wheelchair. Horizontal black lines (20 mm thick) were individually displayed on the board, centred both horizontally and vertically; the lines were lengths of PVC (polyvinylchloride) tape that could be easily attached to (or detached from) the surface of the board. Six line lengths were used: 27, 56, 83, 114, 138 and 166 cm. The patient was positioned 2.44 m from the board, so that the centre of each line fell on his midsagittal plane. Viewing distance is far outside peripersonal space and the visual angles subtended by the line stimuli are roughly the same as those used in step 1. The patient now bisected these lines with a lightweight theatre pointing-light held in the right hand. This pointer projected a thin v-shape cleanly on the screen. T.M. was asked to make a single movement to the centre of the line and then indicate his judgment by depressing a small button on the top of the pointer, thereby producing the illuminated v. Step 3: The apparatus was the same as that used in step 2. The patient (a keen darts player) was now requested to mark the midpoint of each line by throwing a dart from a distance of 2.44 m, which is the normal throwing distance for the game. Step 4: This stage is the same in all respects as step 1. It is a control for any effects of experience or 'learning'. For each step, six trials per line length were given in pseudo-random order, and the horizontal displacement of T.M.'s transections from true centre was measured to the nearest millimetre.

Mean displacements (+ for rightwards, - for leftwards) and standard deviations for each condition are shown in Table 1. Statistical comparisons (paired *t*-test) are based upon absolute (signed) displacements. There is no significant difference between conditions 1 and 4 ($t = 0.085$, degrees of freedom (d.f.) = 5), and hence no 'learning' effect across the experimental session. By contrast, the difference between the results of conditions 1 and 2 is significant ($t = 4.9$, d.f. = 5, $P < 0.01$). Likewise, the other far-space condition (step 3), yields a data set that is significantly more accurate than was found in near space in step 1 ($t = -3.95$, d.f. = 5, $P < 0.01$).

Two final controls were run in separate sessions. Some patients with neglect can bisect rectangles and squares with greater accuracy than lines. In the preceding experiment, the vertical visual angle subtended by the lines in far space was almost double that of the lines in near space. Accordingly, we replicated

steps 1 and 4 with lines of the same length but now 4 mm thick. The results did not differ significantly from those obtained with lines 2 mm thick; variation in the vertical visual angle of the stimuli is therefore not responsible for the observed differences in performance in near and far space. Six weeks after the main experiment we also replicated steps 1 and 4 in a condition where T.M. bisected each line six times in near space with a pointing light held immediately in front of the body (step 5). An analysis of variance with repeated measures, found no significant difference ($f = 8.01$, d.o.f. = 5, 2) between conditions 1, 4 and 5. The differences between performance in near and far space are therefore not due to the response-mode *per se*. There is a significant difference between conditions 2 and 5 ($t = -2.56$, d.o.f. = 5, $P < 0.02$), both of which employ a light-pen.

The data thus show that severe left visual neglect in peripersonal space can coexist with minimal or no neglect in extrapersonal space. Previous failures to discover this dissociation⁷ may indicate the rarity of isolated peripersonal neglect; most patients with persistent left visual neglect have large lesions in the territory of the right middle cerebral artery that may "disrupt more than a single neural system"⁷. Alternatively, these failures may reflect an inadequate choice of experimental task (for example, tests of purely perceptual judgment)⁷. Our extrapersonal tasks require explicit visuo-motor skills that might be expected to map onto an extrapersonal spatial field dedicated to pointing and throwing functions². In future work, it would be pertinent to investigate the pattern of eye movements to near and distant objects; subjects may scan a display differently according to its perceived distance from the viewer.

Previc has recently suggested⁸ that "functional differences between near and far visual space are correlated with their disproportionate representations in the dorsal and ventral divisions of visual association cortex, respectively, and in the magnocellular and parvocellular pathways that project to them."⁸ Our results corroborate the functional distinction, although the extent of T.M.'s lesion is not compatible with the emphasis placed on the role of temporal cortex in attending to far space^{4,8}. The relative preservation of frontal cortex (area 8 in particular) may thus be responsible for T.M.'s accurate performance in extrapersonal space⁹. This extrapolation from studies in monkeys could, however, be questioned.

Some studies of monkey show clear contralateral visual neglect for near space after removal of area 8 (refs 10, 11). When Rizzolatti *et al.*³ reported more severe neglect in far than in near space after ablation of area 8, the stimulus may have been less prominent in far space; the stimuli to which the animals responded may have subtended a much greater visual angle when they were close to the head.

Our behavioural findings extend previous reports of dissociations between personal neglect (in the sense of failure to point on command to a part of one's own body) and failure to explore adequately visual arrays presented within arm's reach¹². Personal neglect (not shown by T.M.) is much less frequent than is neglect of external visual stimuli¹². Above all, the distinction seen in T.M. between reaching and pointing or throwing in far space confirms the teleology of action stressed by Brouchon *et al.*²: "... the ultimate intention of action can determine not only the elaboration and the execution of the movement but also, at least in part, the processing of the perceptual cues associated to the movement." □

7. Pizzamiglio, L. *et al.* *Cortex* **25**, 471–477 (1989).
8. Previc, F. H. *Behav. Brain Sci.* **13**, 519–575 (1990).
9. Rizzolatti, G. & Gallese, V. in *Handbook of Neuropsychology* Vol. 1 (eds Boller, F. & Grafman, J.) 223–246 (Elsevier, Amsterdam, 1988).
10. Kennard, M. A. *Archives of Neurology and Psychiatry* **41**, 1153–1165 (1939).
11. Lattin, R. in *Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 671–691 (MIT Press, Cambridge, Massachusetts, 1982).
12. Bisiach, E. *et al.* *Neuropsychologia* **24**, 759–767 (1986).

ACKNOWLEDGEMENTS. We are grateful to P. Anslow and A. Molyneux for their neuroradiological expertise. This work was supported by the Medical Research Council.

Requirement for nerve growth factor in the development of myelinated nociceptors *in vivo*

Amy M. Ritter, Gary R. Lewin, Norbert E. Kremer & Lorne M. Mendell*

Department of Neurobiology and Behavior, State University of New York, Stony Brook, New York 11794, USA

* To whom correspondence should be addressed

IN adult animals, sensory neurons innervating the skin are phenotypically diverse^{1–3}. We have now investigated whether nerve growth factor (NGF) has a physiological role in the development of this diversity. We gave antisera against NGF to rats from postnatal day 1 (PND 1) to adulthood (5 weeks). We found a virtually complete depletion of high threshold mechanoreceptors conducting in the A δ range (2–13 m s⁻¹) in the sural nerve. This afferent type, normally present in large numbers, appeared to have been replaced by D-hair afferents, sensitive mechanoreceptors which normally are relatively rare. NGF deprivation had this effect only in early postnatal life; treatment from postnatal day 14 to adulthood had no effect. We conclude that the presence of NGF postnatally in skin is necessary for the proper phenotypic development of A δ cutaneous nociceptors.

Rats (Sprague-Dawley) were treated from birth for 5 weeks with antisera against mouse 2.5S NGF (anti-NGF). In the terminal experiments, extracellular single unit recordings were made from 123 A δ (2–13 m s⁻¹) and 175 A β afferents (> 13 m s⁻¹) (Fig. 1a). The physiological properties of A β fibres, such as receptive field size, threshold, modality and conduction velocity, were indistinguishable from those of units in control animals.

By contrast, the population of A δ fibres was radically altered. In controls, the largest proportion (45%) consisted of cutaneous high threshold mechanoreceptors (HTMRs), which encode noxious mechanical events (Fig. 1a) which can lead to pain^{2,4–6}. Low threshold hair afferents, D-hairs^{1,5}, constituted a smaller percentage of the total (26%), and the remaining A δ afferents (29%, see Table 1) were usually high threshold but innervated subcutaneous tissue ('deep'). In animals treated from birth with anti-NGF, cutaneous HTMRs all but disappeared (5–7% of the total), and most (64–72%) of the cutaneous A δ fibres innervated D-hairs (see Table 1). But deep afferents were similar to those in controls both in relative frequency (Fig. 2; Table 1) and in adequate stimulus.

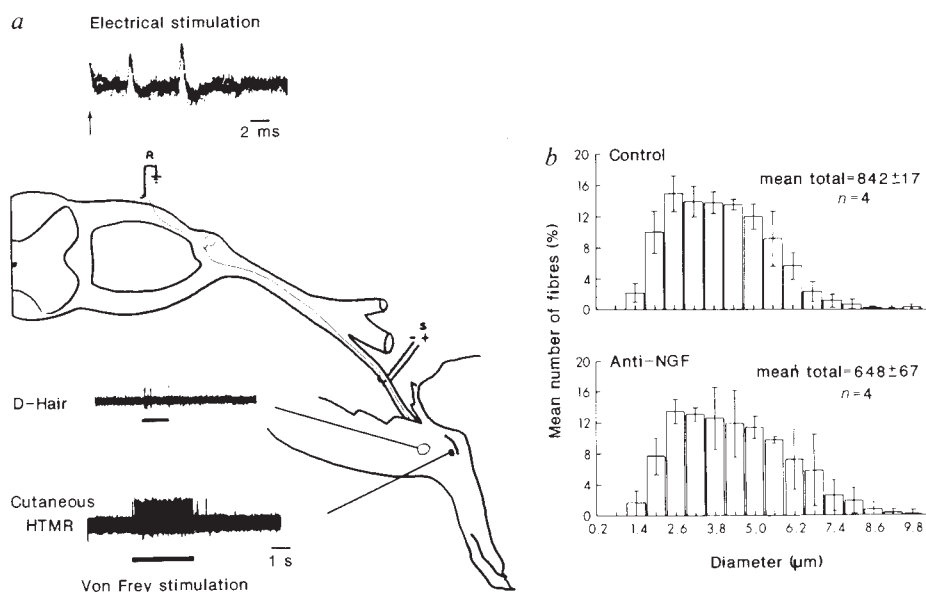
TABLE 1 Physiology of A δ fibres

	Control	Anti-NGF 2–5 weeks	Anti-NGF 0–5 weeks	Anti-NGF +recovery
HTMR	14 (45%)	15 (38%)	1 (5%)	2 (7%)
D-Hair	8 (26%)	12 (31%)	16 (72%)	20 (64%)
Deep	9 (29%)	12 (31%)	5 (23%)	9 (29%)
Total	31 (n=8)	39 (n=3)	22 (n=5)	31 (n=3)

Received 18 January; accepted 21 February 1991.

1. Brain, W. R. *Brain* **64**, 244–272 (1941).
2. Brouchon, M., Joannette, Y. & Samson, M. in *The Biological Foundations of Gestures: Motor and Semantic Aspects* (eds Nespoulous, J.-L., Perron, P. & Lecours, A. R.) 95–107 (Erlbaum, London, 1986).
3. Rizzolatti, G., Matelli, M. & Pavesi, G. *Brain* **106**, 655–673 (1983).
4. Heilman, K. M., Bowers, D. & Shelton, P. *Behav. Brain Sci.* **13**, 552–553 (1990).
5. Wilson, B., Cockburn, J. & Halligan, P. W. *Behavioural Inattention Test* (Thames Valley Test Company, Titchfield, UK, 1987).
6. Halligan, P. W., Manning, L. & Marshall, J. C. *Neuropsychologia* **28**, 1043–1051 (1990).

FIG. 1 *a*, Recordings from sural nerve A δ fibres; *b*, sural nerve myelinated fibre diameter distributions. *a*, Top record, two units driven by electrical stimulation of the sural nerve (SN). Arrow indicates time of nerve stimulation (5 superimposed sweeps). The later (slower conducting) unit, a cutaneous HTMR, is shown (bottom left) in response to suprathreshold Von Frey hair stimulation (8.5-g force). Electrically evoked spikes could be collided with naturally evoked spikes in each case. Middle record shows a D-hair isolated from another strand, demonstrating a typical rapidly adapting discharge (0.0045-g Von Frey hair). D-hairs were physiologically distinct from A β hair follicle afferents; they usually had lower Von Frey thresholds (90% were less than 0.02 g force) than A β hair follicle afferents (70% were between 0.02- and 0.5-g force). Furthermore, most D-hair afferents could be driven by movement of almost every hair within the receptive field, whereas most A β hair afferents were activated by movement of only a few hairs scattered throughout the receptive field. All D-hairs recorded in anti-NGF-treated animals met these two criteria. In rats anaesthetized with urethane (1.25 g kg^{-1}) recordings were made from SN afferents in teased dorsal root filaments¹⁴. We attempted to characterize every well isolated afferent with a latency indicative of A δ conduction velocity; receptive fields were found for 95% of these. Units with no receptive fields were excluded from the analysis. For each functionally single unit isolated, receptive field, adequate stimulus¹⁵, conduction velocity, and Von Frey threshold were noted. *b*, Histological processing of SN and



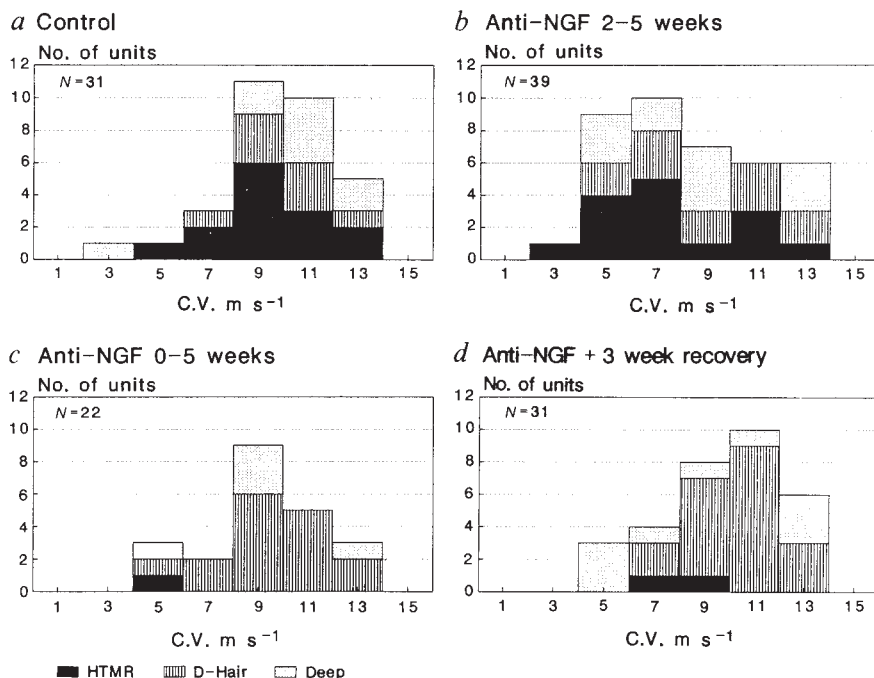
measurement of myelinated fibre diameters have been described⁸. Distributions of myelinated fibre diameters were constructed for nerves from four control and four anti-NGF-treated animals (0–5 weeks). The mean percentage of fibres in each bin (s.d. denoted by bars) was then calculated to produce the data shown in *b*.

These data indicate that cutaneous HTMRs had not merely disappeared, because the ratio of D-hair afferents to deep afferents should have remained about 1:1 rather than the 2:1 ratio found in rats treated from birth with anti-NGF (Table 1 and Fig. 2); this deviation from the expected ratio was significant (χ -squared test, $P < 0.005$). Furthermore, the D-hair afferents were unlikely to be A β hair afferents whose conduction velocity had decreased: their physiological properties were characteristic of D-hairs and thereby very different from those of A β hair follicle afferents (see legend to Fig. 1*a*). Also, A β afferents in anti-NGF-treated rats had normal conduction velocities, as

confirmed by the similarity in distribution of fibre diameters (Fig. 1*b*).

In further experiments, recordings made from 8-week-old animals in which anti-NGF treatment had ceased at 5 weeks showed that the numbers of HTMRs did not recover; most A δ fibres innervated D-hairs (Fig. 2), indicating that the change was permanent. Because the physiology of cutaneous receptors continues to mature in the first 2 weeks of postnatal life⁷, we treated animals with anti-NGF from 2 weeks to 5 weeks before recording from the mature A δ fibres. The proportions of HTMRs, D-hairs and deep units were very similar to controls

FIG. 2 *a–d*, Conduction velocity histograms of A δ units broken down by afferent type. Three groups of afferents are represented: HTMRs, D-hairs, and deep units, which were usually high threshold. Data in each panel are from the following treatment groups: *a*, untreated control animals ($n=8$); *b*, treatment from 2 to 5 weeks of age, ($n=3$); *c*, treatment from birth to 5 weeks of age ($n=5$); and *d*, treatment from birth to 5 weeks of age, then left without treatment for an additional 3 weeks ($n=3$). HTMRs are markedly absent from *c* and *d*, in which anti-NGF treatment was started on the day of birth. No effect is seen in *b*, where treatment was started 2 weeks postnatally. Statistical analysis revealed significant variation for different treatments ($P < 0.001$, χ -squared test), but not between animals of a given treatment group ($P > 0.2$, χ -squared). Antisera against mouse 2.5S NGF¹⁶ were raised in adult rabbits, and were tested for their ability to recognize NGF on western blots and to block NGF-induced neurite outgrowth from PC12 cells¹⁷. Animals were treated with antisera ($5 \mu\text{g l}^{-1}$, subcutaneous) every day for the first week, and then every other day until the end of the treatment. Pre-immune sera were without effect (A.M.R. and L.M., manuscript in preparation).



(Fig. 2, Table 1). Thus manipulation of the afferent population in this way is possible only in the immediate postnatal period.

These results cannot easily be ascribed to sampling bias as units were chosen on the basis of their response to electrical stimulation, and every A δ unit was characterized as to adequate stimulus. The increased branching of dorsal root fibres noted after anti-NGF treatment⁸ could increase the availability of D-hair axons if branching were restricted to these afferents. But an increase larger than the reported 30% would be required to account for the small number of HTMRs sampled.

Because a loss of 17–38% of sensory neurons occurs after neonatal anti-NGF treatment^{8,9}, one might postulate that among myelinated afferents, only HTMRs are dependent on NGF for survival, and in its absence they die. Cross sections of the SN confirmed that on average 23% of the myelinated fibres were missing ($P < 0.005$, unpaired t -test) (Fig. 1b). But if a disproportionate number of the cells that die are A δ HTMRs, there would be a large deficit (about 40%) in small-diameter fibres. In fact the fibre diameter spectra in treated and control rats were not significantly different ($P > 0.6$; nested analysis of variants (ANOVA), $n = 8$; Fig. 1b). These anatomical data, in conjunction with the constant relative frequency of deep afferents (see above), provide strong circumstantial evidence that A δ HTMR afferents were not specifically targeted for death in the absence of NGF.

Alternatively, the absence of NGF could cause presumptive HTMRs to change their phenotype to that of D-hairs. During normal development, afferents destined to innervate hairs first form intra-epidermal projections that persist until around postnatal day 5 after which they withdraw to the dermis¹⁰. But some intra-epidermal endings persist in the adult; these could be the anatomical substrates of A δ nociceptors¹¹. Developing HTMRs could require NGF to stabilize their epidermal projection; in embryonic mice this is where NGF messenger RNA is most concentrated¹². Reduction of NGF concentrations may force fibres to withdraw to the dermis and become D-hair afferents. Thus NGF would be acting indirectly, the phenotypic switch being mediated more directly by dermal factors.

The presence of NGF has usually been thought to regulate the innervation density of certain targets by promoting the survival of only a part of the original innervation¹³. Our findings suggest that this effect may be exerted on a specific class of afferent, the A δ HTMR, either by ensuring their survival or more probably by providing a hospitable environment in which their receptors can form. Thus, NGF may regulate the character, as well as the density, of skin innervation. $\square$

Molecular characterization of single memory B cells

Michael G. McHeyzer-Williams, G. J. V. Nossal & Paul A. Lalor

The Walter and Eliza Hall Institute of Medical Research, Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

PRIMARY antigenic exposure results in an initial antibody response and the T cell-dependent induction of B-cell memory^{1,2}. Memory B-cell differentiation is characterized by somatic hypermutation in antibody variable region genes (V) and selection of B cells expressing high-affinity variants of this antigen receptor^{3–5}. Despite our current understanding of B-cell memory^{6–8}, the origin of memory B cells and the regulation of their differentiation remain elusive. This is largely due to the difficulties in observing and purifying this minor component of the immunized spleen. Further, molecular characterization of memory B cells requires hybridoma formation which restricts analyses to only those clones capable of fusion and does not allow isolation of cells in a normal physiological state. We have therefore developed a unique system which allows isolation and unambiguous enumeration of IgG₁⁺ memory B cells, based on six-parameter flow cytometry, secretion of antibody in clonal cultures and analysis of clonally expressed V genes using the polymerase chain reaction⁹. Here we report that single IgG₁⁺ antigen-binding B cells from an early secondary immune response proliferate in lipopolysaccharide-driven microcultures and produce antigen-specific IgG₁ antibodies. Individual B-cell clones in these cultures express somatically mutated heavy chain V genes, confirming their designation as memory B cells. Although isolated memory B cells undergo extensive proliferation *in vitro*, V gene sequence analysis of their individual progeny shows that further hypermutation does not occur.

Memory B cells generated by T cell-dependent immunization with 4-hydroxy-3-nitrophenylacetyl (NP) respond to a secondary antigenic challenge to produce predominantly IgG₁ antibodies with higher affinities for NP than antibodies in the primary response^{10,11}. The gene combination of λ_1 antibody light chain and V186.2 heavy chain in germ-line configuration are predominantly used by B cells in a primary response^{12,13}. This V gene combination is also expressed by memory B cells involved in the secondary response, but in somatically mutated form¹⁴. Furthermore, the latter cells have a distinct clonal origin from those dominating the primary response¹⁵. The V186.2, λ_1 memory clones use heterogeneous diversity (D) elements with numerous N-sequence insertions, whereas the primary response clones use the germ-line DF116.1 D element with N-sequence insertions only rarely found¹⁵. Also, a proportion of the B-cell clones involved in the secondary response use κ light chain genes with different VH genes¹³.

To generate a population of NP-specific memory B cells, C57Bl/6 mice were primed with 200 μ g NP/keyhole limpet haemocyanin (KLH) and 8 weeks later challenged with 20 μ g soluble NP-KLH. Seven to eight days after secondary immunization, spleens from individual mice were examined using six-parameter flow cytometry for the presence of B cells engaged in the anti-NP response. Propidium iodide and phycoerythrin-conjugated antibodies to IgM, MAC-1 (CD 11b) and GR-1 were used collectively to stain dead cells, IgM⁺ B cells and macrophages which were then excluded. Among the IgM⁺ lymphocyte population, NP-specific B cells were identified by expression of IgG₁ and ability to bind to NP coupled to the fluorescent phycobiliprotein, allophycocyanin (Fig. 1). A distinct population of IgG₁⁺IgM⁺ cells was detected, consisting of 0.2% of the whole spleen. Of these cells, 60% (0.12% of splenocytes, or 1.6×10^5 cells per spleen) bound NP. Most (60%) of these IgG₁⁺IgM⁺NP⁺ cells used the λ light chain. Hence, the population of IgG₁⁺IgM⁺NP⁺ cells represented 0.07% of splenocytes,

Received 6 November 1990; accepted 12 February 1991.

- Burgess, P. R. & Perl, E. R. in *Handbook of Sensory Physiology* (ed. Iggo A.) **2**, 29–78 (Springer, Berlin, 1973).
- Burgess, P. R. & Perl, E. R. *J. Physiol.* **190**, 541–562 (1967).
- Koerber, H. R., Druzinsky, R. E. & Mendell, L. M. *J. Neurophysiol.* **60**, 1584–1595 (1988).
- Lynn, B. & Shakhnabeh, J. *Neurosci. Lett.* **85**, 71–76 (1988).
- Lynn, B. & Carpenter, S. E. *Brain Res.* **238**, 29–43 (1982).
- Handwerker, H. O., Adriaensson, H., Gybels, J. & Van Hees, J. in *Development, Organization, and Processing in Somatosensory Pathways* (eds Rowe, M. J. & Willis, W. D.) 289–297 (Alan Liss, New York, 1985).
- Fitzgerald, M. J. *Physiol.* **333**, 79–92 (1987).
- Hulsebosch, C. E., Perez-Polo, J. R. & Coggeshall, R. E. *J. comp. Neurol.* **259**, 445–451 (1987).
- Yip, H. K., Rich, K. M., Lampe, P. A. & Johnson, E. M. Jr. *J. Neurosci.* **4**, 2986–2992 (1984).
- Fitzgerald, M. J. *J. comp. Neurol.* **126**, 37–42 (1966).
- Kruger, L., Perl, E. R. & Sedivec, M. J. (1981) *J. comp. Neurol.* **198**, 137–154 (1981).
- Davies, A. M. *et al. Nature* **326**, 353–358 (1987).
- Barde, Y.-A. *Neuron* **2**, 1525–1534 (1989).
- Lewin, G. R. & McMahon, S. B. *J. Physiol.* **412**, 26P (1989).
- Horch, K. W., Tuckett, R. P. & Burgess, P. R. *J. invest. Dermatol.* **69**, 75–82 (1977).
- Mobley, W. C., Schenker, A. & Shooter, E. M. *Biochemistry* **15**, 5543–5552 (1976).
- Greene, L. A. *Brain Res.* **133**, 350–353 (1977).

ACKNOWLEDGEMENTS. We thank S. Halegoua for use of facilities for isolating NGF and preparing the anti-sera, A. Bushnell for technical assistance, N. Mendell for statistical analysis, and M. Fitzgerald and A. Light for helpful comments. Supported by NIH.

or 10^5 cells per spleen. About 70% of these antigen-specific IgG_1^+ cells showed increased forward and orthogonal light scattering properties compared to the IgM^+ B cells in the spleen, consistent with activation.

For subsequent functional and molecular characterization we concentrated on the λ -bearing IgG_1^+ antigen-specific B cells because of their known association with use of the *V186.2* gene with λ in the NP response. Single cells from this putative memory B-cell population were sorted into microculture wells and stimulated to proliferate and secrete antibody by addition of lipopolysaccharide, recombinant interleukins 2, 4 and 5 and 3T3 fibroblast filler cells^{16,17}. Under these conditions, one in six of the $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ B cells produced high levels of $\lambda^+\text{IgG}_1$ antibody as detected by enzyme-linked immunosorbent assay (ELISA); a minimum of 1 ng IgG_1 was produced per clone and more than 10 ng in 50% of clones. The IgG_1 produced by 95% of the clones bound NP in ELISA (Table 1). The culture system also allowed for B-cell clones to be visually monitored in the microculture wells. After six days, one in four of the clones that produced antigen-specific IgG_1 had up to 250 progeny (lower level of detection was 40 progeny cells). Overall, the cloning efficiency of the putative memory B cells for IgG_1 production was comparable to that of IgM^+ B cells selected by flow flow cytometry and cultured under the same conditions (unpub-

lished observations). The $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ B cells gave rise to a slightly lower frequency of $\lambda^+\text{IgG}_1$ antibody-producing clones (1/8), of which only 15% bound NP (Table 1).

Flow cytometry has previously been used to isolate populations of IgM^- , IgD^- antigen-specific memory B cells^{18,6}. Here, however, we have directly identified their *in vivo* precursors by expression of surface IgG_1 , and have clonally expanded them in the absence of T cells or antigen. Our functional analyses of single cells strongly suggest that the $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ phenotype defines a pure population of B cells involved in the secondary NP response. There was, however, some evidence of heterogeneity within this population (Fig. 1). Cells identified as the highest NP-binders showed lower *in vitro* cloning ability than the total $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ population (Table 1), perhaps reflecting different physiological states within this population of putative memory B cells.

Analysis of VH gene somatic mutation was used to confirm the designation of the $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ population as memory B cells. A total of 50 different IgG_1 , λ^+ anti-NP producing clones, derived from two separate mice, was analysed for VH expression using polymerase chain reaction (PCR) primers specific for *V186.2* heavy chain gene rearranged to IgG_1 . A total of 26/50 clones gave a DNA fragment of the expected size after two separate rounds of PCR. The remaining clones that were negative for *V186.2* may have used different VH genes in association with λ to produce anti-NP antibodies. Alternatively, these clones may have used the *V186.2* gene, but the levels of initial messenger RNA or complementary DNA prepared from these clones may have been limiting in the PCR. Sequencing showed that 20 of the 26 positive clones used the *V186.2* gene in mutated form, with between 3 and 11 mutations per gene. A representative population is presented in Fig. 2. The presence of somatic mutations in the expressed *V186.2* genes confirms that the $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ B cells identified and quantified by flow cytometry contain memory B cells.

Although 4 of the 26 clones used the *V186.2* gene in germ-line configuration, all clones used a different pattern of junctional diversity indicating they had arisen independently during the *in vivo* response. Furthermore, all clones used heterogeneous D elements with extensive N-sequence insertions which may be characteristic of secondary clones in the NP response^{13,15}. Two of the 26 clones analysed used the *V23* gene, which is highly homologous to *V186.2* (ref. 12), with each clone again showing mutations. Nine clones were also analysed from the sorted $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ population which produced $\lambda^+\text{IgG}_1$ antibodies that did not bind NP by ELISA. Only one clone gave

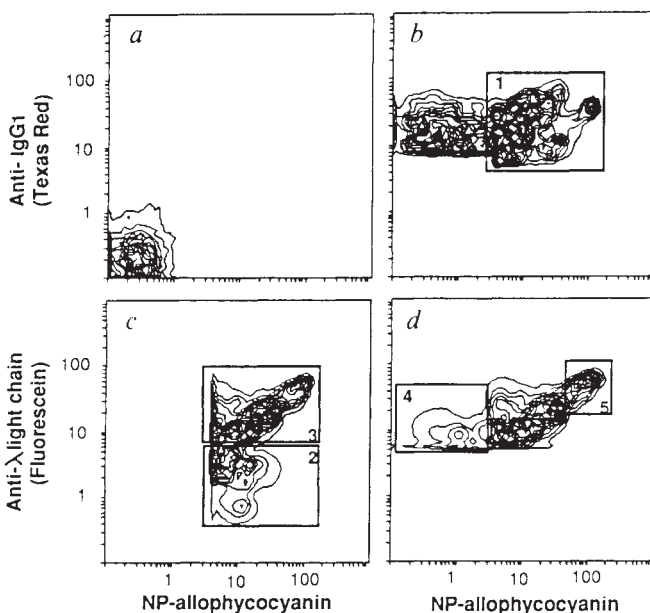


FIG. 1 Flow cytometric selection of a memory B cell population. *a*, NP binding and IgG_1 expression of lymphocytes excluding IgM^+ , MAC-1^+ and GR-1^+ cells (50% of splenocytes). *b*, NP binding to $\text{IgG}_1^+\text{IgM}^-$ B cells. These cells comprise 0.4% of the cells in *a*. Box 1 shows the $\text{IgG}_1^+\text{IgM}^-\text{NP}^+$ population (57% of $\text{IgG}_1^+\text{IgM}^-$ cells). *c*, The λ light-chain expression of $\text{IgG}_1^+\text{IgM}^-\text{NP}^+$ cells shown in box 1. Box 2 represents the λ^- fraction (35%) and box 3 the λ^+ fraction (65%). *d*, NP binding on all $\text{IgG}_1^+\text{IgM}^-\lambda^+$ cells. This shows that 90% of $\text{IgG}_1^+\text{IgM}^-\lambda^+$ B cells bind NP. Box 4 shows the $\text{IgG}_1^+\text{IgM}^-\lambda^+\text{NP}^+$ cells; Box 5 represents the NP^+ cells in the putative memory population. METHODS. C57BL/6 mice were immunized intraperitoneally (i.p.) with 200 μg alum-adsorbed NP-KLH and 10^9 killed *Bordetella pertussis*, rested for 8 weeks, challenged with 20 μg soluble NP-KLH (i.p.) and sacrificed 7–8 days later. Spleen cells (1.3×10^9 per mouse) were isolated and stained with biotinylated goat anti-mouse IgG_1 (Southern Biotechnology Associates, Birmingham, Alabama), followed by Texas Red-avidin, fluoresceinated-JC5 (anti- λ light chain) (provided by L. A. Herzenberg), NP-coupled allophycocyanin and phycoerythrin conjugates of 331.12 (anti-mouse IgM monoclonal antibody), 8C5 (anti-GR1 monoclonal antibody) and MAC-1 (CD 11b). Dead cells were excluded by staining with propidium iodide. All antibodies and staining techniques have been described²⁶. Cells were sorted using a FACStar Plus (Becton-Dickinson) and an automatic cell deposition unit. All profiles show data as probability curves (6.7% spacing).

TABLE 1 Response *in vitro** from FACS-selected memory B cells

Phenotype‡	Cells per spleen	Number of positive clones per 1,000 input cells†			
		Total IgM	Total IgG_1	λ^+ IgG_1	Anti-NP IgG_1
Unfractionated	1.3×10^8	160	200	5	3
$\text{IgM}^+\text{IgG}_1^-\lambda^+\text{NP}^+$	9.0×10^4	7	170	160	160
$\text{IgM}^+\text{IgG}_1^-\lambda^+\text{NP}^-$	1.4×10^4	0	125	110	17
$\text{IgM}^+\text{IgG}_1^-\lambda^+\text{NP}^{++}$	1.4×10^4	0	30	30	30

* Cells were cultured with 2,500 BALB/c 3T3 fibroblasts per well, 40 μg per ml lipopolysaccharide, 50 U per ml recombinant interleukin 4 (rIL-4), 20 U per ml rIL-2, 2% rIL-5-containing supernatant (v/v) for 8 days at 37 °C in 10% CO_2 and air¹⁷.

† Frequencies were calculated from data for 288 cells from each selected population, deposited into individual wells of 96-well microculture plates or by limiting dilution analysis on unfractionated cells. The isotype, light chain use and NP-binding ability of secreted antibodies were determined in sensitive ELISA, as described^{17,25}. One representative experiment is shown.

‡ Phenotype corresponds to the selected populations as described in Fig. 1; unfractionated spleen cells were shown to contain 50% IgM^+ B cells in these experiments, after removal of erythroid cells by NH_4Cl lysis.

CLONE	CDR1	CDR2	D	J _H
	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18		
V186.2	QPAQLASKLKGTSYWMHVLQGREIRINSSGGTKWNESKATLTVKPS	STAGSSEYARY		
1	---	---	?	2
2	---	---	S	3
3	---	---	F	4
4	---	---	?	2
5	---	---	F	1
6	---	---	S	2
7	---	---	S	2
8	---	---	F	2
9	---	---	F	2
10	---	---	?	1
11	---	---	F	1
12	---	---	F	1
13	---	---	S	2
14	---	---	S	2
15	---	---	F	2
16	---	---	?	2
V23	---	---		
17	---	---	S	1
18	---	---	S	2

FIG. 2 Somatic mutation in V_H genes of memory B-cell clones. The predicted amino-acid sequences of V_H chains derived from 18 distinct B cell clones are compared with the germ-line V186.2. Numbers above sequence indicate amino-acid position, CDR represents the complementarity determining regions 1 and 2. The assignment of D family (more than six nucleotide similarities with not more than one mismatch; F, DFL16; S, SP2) and J_H use are listed. Replacement mutations are represented by the amino-acid changes; asterisk, a silent mutation; dashed line, identity. Clones 1 to 15, 17 and 18 were selected from clonal cultures of IgG₁ IgM⁺ λ⁺NP⁺ cells, which produced IgG₁ anti-NP antibody detected by ELISA; clone 16 was selected as IgG₁ IgM⁺ λ⁺NP⁻ and produced λ⁺IgG₁ antibody which did not bind NP. Clones 17 and 18 were more closely related to V23 (ref. 12) than to V186.2.

METHODS. Total RNA was extracted from individual clones²⁷, cDNA synthe-

sized using oligo(dT) primed reverse transcription²⁸ and PCR performed using ampli-Taq (Perkin Elmer Cetus) (1 unit per 30 μl reaction) with standard buffer containing 2 mM MgCl₂, 0.2 mM mixed dNTPs and 1 μM oligomer primers⁹. In the first set of PCR amplification, primers used were as follows: GGATGACTCATCCAGGGTCACCATGGAGT (for CH_γ1); GCTGTATCATGCTCTCTTG (for V186.2 leader sequence). Forty cycles were performed: melting at 94 °C for 45 s; annealing at 55 °C for 60 s; elongation at 70 °C for 90 s. Two microlitres from the first reaction were used for a second set of 40 cycles in identical conditions using oligomers internal to the first set: CCAGGGGCCAGTGGATAGAC (for CH_γ1); GGTGTCCACTCCAGGTCCA (for V186.2). The PCR products were purified by electrophoresis through low melt agarose gels and isolated as described²⁸. Direct double-stranded DNA sequencing was performed by the dideoxy-chain-termination method²⁹ with modifications³⁰, using the primers from the second set of PCR.

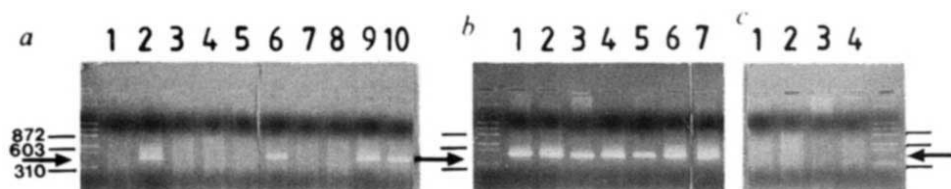
rise to a DNA fragment containing the V186.2 gene (clone 16 in Fig. 2). The single mutation in position 52 in this clone may have resulted in loss of specificity, although specificity loss due to mutation in the λ gene has not been excluded.

To prove that the sequence changes observed in the B-cell clones were a result of *in vivo* hypermutation and were not a PCR artefact, we performed PCR and sequence analysis on individual, micromanipulated progeny cells from clonal cultures. Individual cells were micromanipulated from clones containing at least 250 cells after 6 days of culture and the remaining cells were pooled for determination of V186.2 gene use by that clone. Forty per cent of individual cells from two clones (clones 5 and 16) gave rise to a DNA fragment of the expected size

(Figs 2 and 3). All six fragments prepared from separate cells of clone 5 contained an identical DNA sequence showing three replacement mutations within the V186.2 gene. Eleven separate samples from clone 16 also contained an identical sequence with a single point mutation at position 52 in V186.2. Therefore, the mutations observed in the analysis of different clones are the result of hypermutation *in vivo*.

Furthermore, these results demonstrate the absence of continued hypermutation within memory B cells during the period of extensive *in vitro* proliferation. If hypermutation were to occur at the reported frequency of 10⁻³ per base pair per division *in vitro*¹⁹, the probability of finding six cells in the same clone after eight cell divisions using the same sequence is of the order

FIG. 3 V_H gene PCR products from individual progeny cells. An agarose gel stained with ethidium bromide is presented with the arrow indicating the expected DNA fragment size of 380–390 bp (depending on extent of N-sequence insertion). a, Lanes 1–10 represent the product from RNA of micromanipulated single cells. b, Lanes 1–6 represent RNA from 2, 4, 8, 16, 32 and 64 micromanipulated cells respectively; lane 7 is the product from the cells remaining in the well (>100). c, Negative controls; lane 1, no cell for



RNA extraction; lane 2, no RNA for cDNA synthesis; lane 3, no cDNA for first set of PCR; lane 4, no PCR product for second set of PCR.

of 10^{-11} . Similarly, proliferation of memory B cells during *in vivo* expansion is insufficient to drive hypermutation²⁰, leading to models in which hypermutation is not merely a consequence of memory B-cell proliferation^{14,21,22}. It is intriguing to consider whether an *in vitro* culture system could be developed which would allow V gene hypermutation in isolated memory B cells.

We have identified a phenotypically discrete, pure population of memory B cells during an immune response, cloned them *in vitro* in the absence of T cells or antigen and molecularly

characterized expressed V genes in single progeny from these clones. Memory B cells are isolated immediately *ex vivo* in a normal physiological state, without requirement for hybridoma formation and its associated limitations. With this unique system we have begun to track IgG⁺, NP-specific B cells back to their point of emergence in a primary response (P.A.L., G.J.V.N. and M.G.M.-W., manuscript in preparation). We intend to establish when the developing memory B cells first undergo mutation, which may be as early as the fifth day of the primary response^{23,24}. □

Received 22 November 1990; accepted 4 February 1991.

- MacLennan, I. C. M., Liu, Y. J., Oldfield, S., Zhang, J. & Lane, P. J. L. *Curr. Topics Microbiol. Immun.* **159**, 37–63 (1990).
- Gray, D. & Leanderson, T. *Curr. Topics Microbiol. Immun.* **159**, 1–17 (1990).
- Weigert, M., Cesari, I. M., Yonkovich, S. J. & Cohn, M. *Nature* **228**, 1045–1048 (1970).
- Berek, C. & Milstein, C. *Immunol. Rev.* **105**, 5–26 (1988).
- Kocks, C. & Rajewsky, K. A. *Rev. Immun.* **7**, 537–559 (1989).
- Schittke, B. & Rajewsky, K. *Nature* **346**, 749–751 (1990).
- Gray, D. & Skarvall, H. *Nature* **336**, 70–73 (1988).
- Linton, P.-J., Decker, D. J. & Klinman, N. R. *Cell* **59**, 1049–1059 (1989).
- Saiki, R. A. *et al. Science* **239**, 487–491 (1988).
- Imanishi, T. & Mäkelä, O. *Eur. J. Immun.* **3**, 323–330 (1973).
- Reth, M., Hämmerling, G. J. & Rajewsky, K. *Eur. J. Immun.* **8**, 393–400 (1978).
- Bothwell, A. L. M. *et al. Cell* **24**, 625–637 (1981).
- Allen, D. *et al. Immunol. Rev.* **96**, 5–22 (1987).
- Rajewsky, K., Förster, I. & Cumano, A. *Science* **238**, 1088–1094 (1987).
- Cumano, A. & Rajewsky, K. *EMBO J.* **5**, 2459–2468 (1986).
- Pike, B. L., Vaux, D. L., Clarke-Lewis, I., Schrader, J. W. & Nossal, G. J. V. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6350–6354 (1982).
- McHeyzer-Williams, M. G. *Eur. J. Immun.* **19**, 2025–2030 (1989).

- Hayakawa, K., Ishii, R., Katsuhiko, Y., Kishimoto, T. & Hardy, R. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1379–1383 (1987).
- McKean, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 3180–3184 (1984).
- Siekevitz, M., Kocks, C., Rajewsky, K. & Dildrop, R. *Cell* **48**, 757–770 (1987).
- Manser, T. *Immunol. Today* **11**, 305–308 (1990).
- Lebecque, S. G. & Gearhart, P. J. *J. exp. Med.* (in the press).
- Levy, N. S., Malipiero, U. V., Lebecque, S. G. & Gearhart, P. J. *J. exp. Med.* **169**, 2007–2019 (1989).
- Tao, W. & Bothwell, A. L. M. *J. Immun.* **145**, 3216–3222 (1990).
- Lalor, P. A. & Morahan, G. *Eur. J. Immun.* **20**, 485–492 (1990).
- Lalor, P. A., Stall, A. M., Adams, S. & Herzenberg, L. A. *Eur. J. Immun.* **19**, 501–506 (1989).
- Chomczynski, P. & Sacchi, N. *Analyst. Biochem.* **162**, 156–159 (1987).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Sanger, F., Miklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Winship, P. R. *Nucleic Acids Res.* **17**, 1266 (1989).

ACKNOWLEDGEMENTS. We thank D. Kemp, S. Foote, A. Troutt and A. Kelso for helpful advice and critical review of manuscript. This work was supported by the National Health and Medical Research Council (Canberra, Australia) and the National Institute for Allergy and Infectious Diseases (Bethesda, Maryland, USA).

Subcellular patterns of calcium release determined by G protein-specific residues of muscarinic receptors

James Lechleiter*, Steven Girard*, David Clapham* & Ernest Peralta†

* Department of Pharmacology, Mayo Foundation, 200 First Street SW, Rochester, Minnesota 55905, USA

† Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

CALCIUM release from intracellular stores is a point of convergence for a variety of receptors involved in cell signalling¹. Consequently, the mechanism(s) by which cells differentiate between individual receptor signals is central to transmembrane communication^{2–5}. There are significant differences in timing and magnitude of Ca²⁺ release stimulated by the m2 and m3 muscarinic acetylcholine receptors^{6–11}. The m2 receptors couple to a pertussis toxin-sensitive G protein to activate phosphatidyl inositol hydrolysis weakly and to stimulate small, delayed and oscillatory chloride currents. In contrast, m3 receptors potentially activate phosphatidyl inositol hydrolysis and stimulate large, rapid and transient chloride currents by a pertussis toxin-insensitive G protein pathway. Using confocal microscopy, we now show that the m2- and m3-coupled Ca²⁺ release pathways can also be spatially distinguished. At submaximal acetylcholine concentrations, both receptors stimulated pulses of Ca²⁺ release from discrete foci in random, periodic and frequently bursting patterns of activity. But maximal stimulation of m2 receptors increased the number of focal release sites, whereas m3 receptors invariably evoked a Ca²⁺ wave propagating rapidly just beneath the plasma membrane surface. Analysis of pertussis toxin sensitivity and hybrid m2–m3 muscarinic acetylcholine receptors confirmed that these Ca²⁺ release patterns represent distinct cell signalling pathways.

Individual *Xenopus* oocytes were assayed 48 h after injection of SP6-generated RNA transcripts encoding m2, m3 or hybrid

muscarinic acetylcholine receptors (mAChRs). Each oocyte was also injected with the Ca²⁺ dye indicator fluo-3 (50 nl of 1 mM solution) 30–120 min before electrophysiological recording and Ca²⁺ imaging analysis. The Ca²⁺ response of an oocyte expressing m2 mAChRs was measured over time from an optical slice at the plasma membrane surface (Fig. 1). Application of 1 μ M acetylcholine (ACh) produced no detectable response. However, a 50 μ M application of ACh induced a Ca²⁺ increase which was characteristically very slow (>20 s to onset), oscillatory and focal (Fig. 1b and c). Heterogeneity of subcellular Ca²⁺ release was typical of the m2-coupled pathway (Table 1). The temporal change in Ca²⁺ is shown by a volume rendition of consecutive images of the same optical slice recorded at 1-s intervals (Fig. 1d). Below this volume of Ca²⁺ activity, the mean fluorescent intensity of each optical slice is plotted (Fig. 1e). Activation of Ca²⁺-dependent Cl[−] currents (*I*_{Ca,Cl}) correlated with changes in intracellular Ca²⁺ (Fig. 1f) and was unaffected by withdrawal of extracellular Ca²⁺. In many oocytes, however, an electrical response was seen without an accompanying increase in Ca²⁺. We attributed this to the focal nature of Ca²⁺ release and to the fact that over 96% of the egg was outside our optical slice (assuming an average oocyte diameter of 1 mm). These observations underscored the unavoidable loss of cell signalling data when relying on current measurements alone. Electrophysiological responses represented an integration of Ca²⁺ release over the entire surface of the cell and contained no spatial information. Deep cytoplasmic signals may not be detectable as Ca²⁺ may be buffered before diffusion to the plasma membrane surface. Control oocytes injected with water consistently failed to respond to ACh (Table 1; see also ref. 6).

The ACh-induced responses of oocytes expressing m3 mAChRs were markedly different compared with the m2-mediated pathway. First, the m3-mediated Ca²⁺ response to agonist was large, rapid and usually transient (Fig. 2). Second, the characteristic spatial pattern of maximal Ca²⁺ release was invariably a propagating wave of fluorescence (Fig. 2a; Table 1) which eventually enveloped the entire oocyte (Fig. 2b). The speed of Ca²⁺ wave propagation ranged from 10–30 μ m s^{−1}. Third, maximal stimulation of m3 responses was observed at an ACh concentration of 1 μ M; this dose of agonist was rarely sufficient to evoke detectable responses from m2-injected oocytes.

Subsequent applications of 50 μM ACh consistently failed to elicit additional responses (Fig. 2c and d), presumably because of depletion of Ca^{2+} stores and/or to receptor desensitization. A rapid and transient $I_{\text{Ca,Cl}}$ was also induced by ACh application. Interestingly, the peak current frequently occurred before the appearance of the Ca^{2+} wave in our optical slice and always decayed more rapidly to baseline than did the mean Ca^{2+} levels (Fig. 2e; see also ref. 12). The focal nature of Ca^{2+} release observed for the m2 pathway could also be induced in some oocytes expressing m3 mAChRs by applying lower concentrations of ACh (10 nM) (Fig. 2f). In contrast to the m2-mediated pathway, a subsequent increase in the ACh concentration (1 μM) to m3-expressing oocytes invariably induced the characteristic propagating Ca^{2+} wave (Fig. 2g and h). At intermediate concentrations of ACh (100 nM), oocytes usually exhibited Ca^{2+} waves of reduced intensity when compared with higher concentrations; a subsequent application of 1 μM ACh produced a second, more intense wave of Ca^{2+} propagation. On several occasions, a 100 nM dose of ACh was sufficient to produce the maximal response, since additional increases in agonist failed to elicit additional Ca^{2+} release. Finally, the m3-induced $I_{\text{Ca,Cl}}$ was temporally correlated with increases in Ca^{2+} and occurred in the absence of extracellular Ca^{2+} . On occasion, $I_{\text{Ca,Cl}}$ activation preceded or followed the Ca^{2+} profile and suggested that Ca^{2+} release occurred outside of the confocal imaging field and/or distal to the plasma membrane surface (Fig. 2i and j).

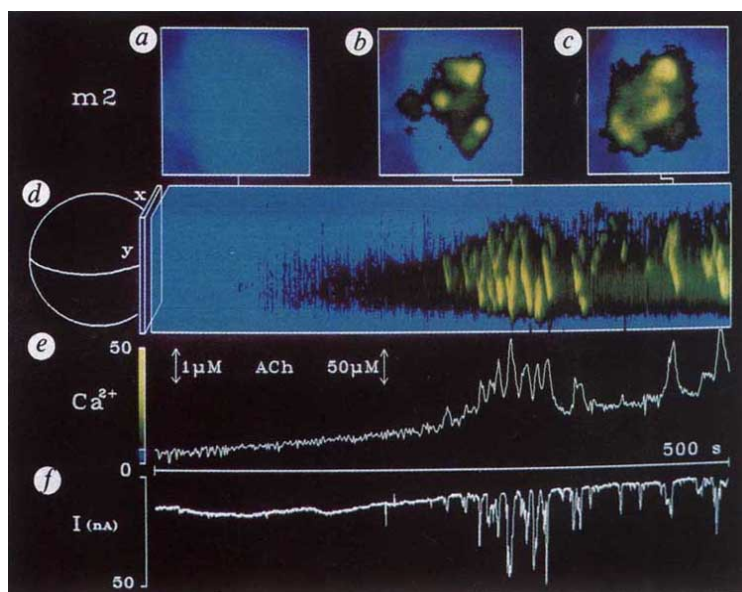
The spatial and temporal patterns of Ca^{2+} release were also examined in oocytes injected with the two hybrid mAChRs, Hy12 and Hy19 (ref. 6). Hy12 is an m3 receptor hybrid in which

the first 21 N-terminal amino acids of the large cytoplasmic loop adjacent to the fifth transmembrane segment have been exchanged with the corresponding m2 residues (Table 1). ACh stimulation of Hy12 expressing-oocytes resulted in Ca^{2+} release patterns that were indistinguishable from m2-mediated responses (Fig. 3A; Table 1). Multiple focal Ca^{2+} release sites were observed at all ACh concentrations larger than 1 μM (Fig. 3A, b–e). Furthermore, the magnitude of intracellular Ca^{2+} release increased with progressively larger doses of ACh (Fig. 3A, f), as did stimulation of $I_{\text{Ca,Cl}}$ (Fig. 3A, g). Note that in the oocyte depicted in Fig. 3A, a burst of Ca^{2+} activity occurred which was similar to that observed with m2-expressing oocytes (Fig. 1d and e).

Hy19 is an m2 receptor hybrid in which only nine cytoplasmic residues proximal to the fifth transmembrane segment have been exchanged for the corresponding m3 amino acids (Table 1). Although 457 out of 466 amino acids of Hy19 are derived from the m2 mAChR subtype, this hybrid receptor induced Ca^{2+} release patterns that were indistinguishable from m3-expressing oocytes (Fig. 3B; Table 1). Hy19 responded maximally at 1 μM ACh and produced the characteristic m3-like wave pattern of Ca^{2+} release (Fig. 3B, a–d). The speed of wave propagation ranged from 10–20 $\mu\text{m s}^{-1}$. Similarly, the rapid $I_{\text{Ca,Cl}}$ evoked by Hy19 was very similar to the electrical responses recorded in oocytes expressing m3 receptors. In this oocyte the mean Ca^{2+} concentration did not decrease over the course of the experiment (Fig. 3B, e), and this presumably activated the relatively high level of plateau current. An elevated plateau phase of Ca^{2+} release was also present in many of the oocytes expressing m3 mAChRs (data not shown).

FIG. 1 ACh-induced Ca^{2+} release in oocytes expressing m2 mAChRs. a–c, Spatial patterns of Ca^{2+} release for single optical slices recorded at 0, 304 and 449 s, respectively. ACh was added at 20 s (1 μM) and 200 s (50 μM). Resting Ca^{2+} is depicted blue on a scale ranging from 0–20 intensity units. Ca^{2+} increases are depicted green–yellow on the same scale ranging from 20 (no change) to 50 intensity units. All Ca^{2+} concentrations depicted are absolute intensity units, ranging from 0–255, because fluo-3 (Molecular Probes) is a non-ratioing Ca^{2+} indicator dye and does not provide absolute Ca^{2+} concentrations. An *in vitro* calibration (see below) placed 0–20 intensity units between 10 and 40 nM with the maximum intensity of 255 corresponding to 300 nM. d, Spatio-temporal nature of Ca^{2+} release. Ca^{2+} activity was rendered as a volume by sequentially stacking 500 images, captured at 1 s intervals, of a single optical slice in the x–y plane of the oocyte. Individual optical slices (a–c) are shown with right edge into the plane of the paper and with the left edge out. Depth is created by displaying only the brightest pixel along the readers line of sight. Thus, the pigment half of the oocyte (note the dark lane in (a) on the left of the resting Ca^{2+} image) is transparent in the volume rendition owing to its relative low intensity pixel values. e, Mean image intensity. Only Ca^{2+} increases are plotted. The green–yellow scale bar begins at the level at which significant Ca^{2+} increases appear in image (d). The blue scale on the bottom left axis illustrates the compressed range of resting Ca^{2+} levels and refers only to the images a–d. f, $I_{\text{Ca,Cl}}$ recorded simultaneously during experiment. Oocyte was voltage clamped at –70 mV.

METHODS. Images (650 $\times$ 650 $\times$ 40 μm , 128 $\times$ 128 pixels, 256 intensity levels) were low pass filtered (9 $\times$ 9 convolution mask) for analysis. The first 20 optical slices were averaged for the resting Ca^{2+} estimate, because the mean background noise on our confocal microscope varied from one to two intensity units per image but did not vary per 20-image average. This frame-to-frame variation in intensity was uniformly distributed throughout individual images and did not affect the spatial information within a single image. No change in resting Ca^{2+} was observed without the addition of ACh. The ACh-induced Ca^{2+} increases were obtained by subtracting the resting Ca^{2+} image from each of the consecutive optical slices. Changes of less than five intensity units were plotted in the mean intensity graph but omitted from images. This did not affect the peak magnitude of Ca^{2+} increases, but



highlighted the patterns of Ca^{2+} release. For final display, images of increased Ca^{2+} levels were overlaid on the image of the resting Ca^{2+} level. At the image locations where green–yellow and blue pixels overlapped, only Ca^{2+} increases are displayed. Thus, in volume renditions the brightest pixel algorithm preferentially displayed the spatial content of Ca^{2+} increases over the spatial content of resting Ca^{2+} . All images were collected with a Lasersharp MRC-600 Bio-Rad confocal box adapted to an IM35 Zeiss microscope (10 $\times$ objective; Zeiss F10 achromat objective, 0.25 NA) and analysed using ANALYZE software (Mayo Foundation) on a Silicon Graphics, Personal Iris computer. The confocal detector aperture was set at the largest opening and the photomultiplier gain at the largest setting to maximize the fluorescence signal. Ca^{2+} concentrations were calibrated *in vitro*^{14,15}. The optical thickness was estimated using the tilted reflector method (Bio-Rad technical bulletin). Experimental conditions and oocyte procedures were as described previously⁶.

FIG. 2 ACh-induced Ca^{2+} release in oocytes expressing m3 mAChRs. *a, b*, Spatial patterns of Ca^{2+} release for single optical slice recorded at 56 and 92 s. Ca^{2+} increases range from 20–230 intensity units. ACh was added at 20 s ($1\ \mu\text{M}$) and at 200 s ($50\ \mu\text{M}$). *c*, Spatio-temporal pattern of the m3-mediated Ca^{2+} release. Volume rendered as described in Fig. 1. *d*, Mean image intensity. *e*, $I_{\text{Ca,Cl}}$ recorded during experiment. Note that $50\ \mu\text{M}$ ACh produces no additional release of Ca^{2+} or stimulation of current. *f*, Multiple focal Ca^{2+} release sites present in a second oocyte exposed to $10\ \text{nM}$ ACh at 10 s. *g*, $1\ \mu\text{M}$ ACh at 280 s induced Ca^{2+} wave. *h*, Volume rendition of Ca^{2+} release. Ca^{2+} increases range from 20–150 intensity units. *i, j*, Mean image intensity and $I_{\text{Ca,Cl}}$ during experiment.

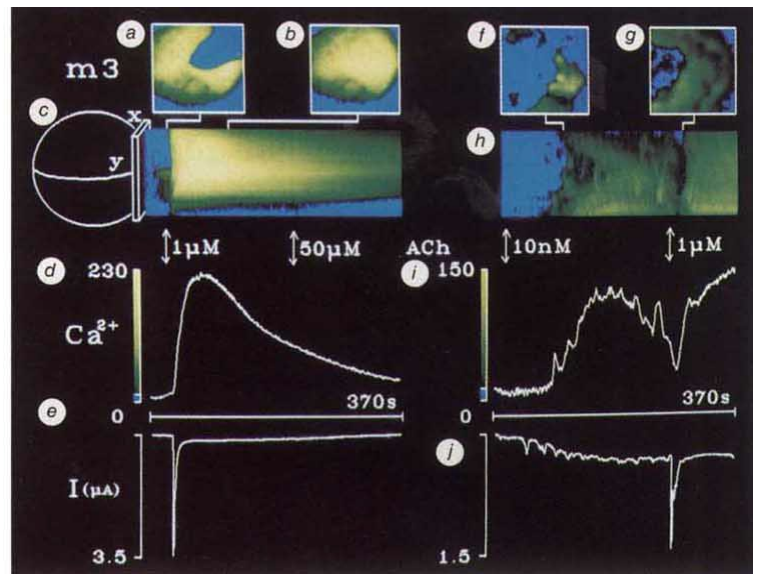
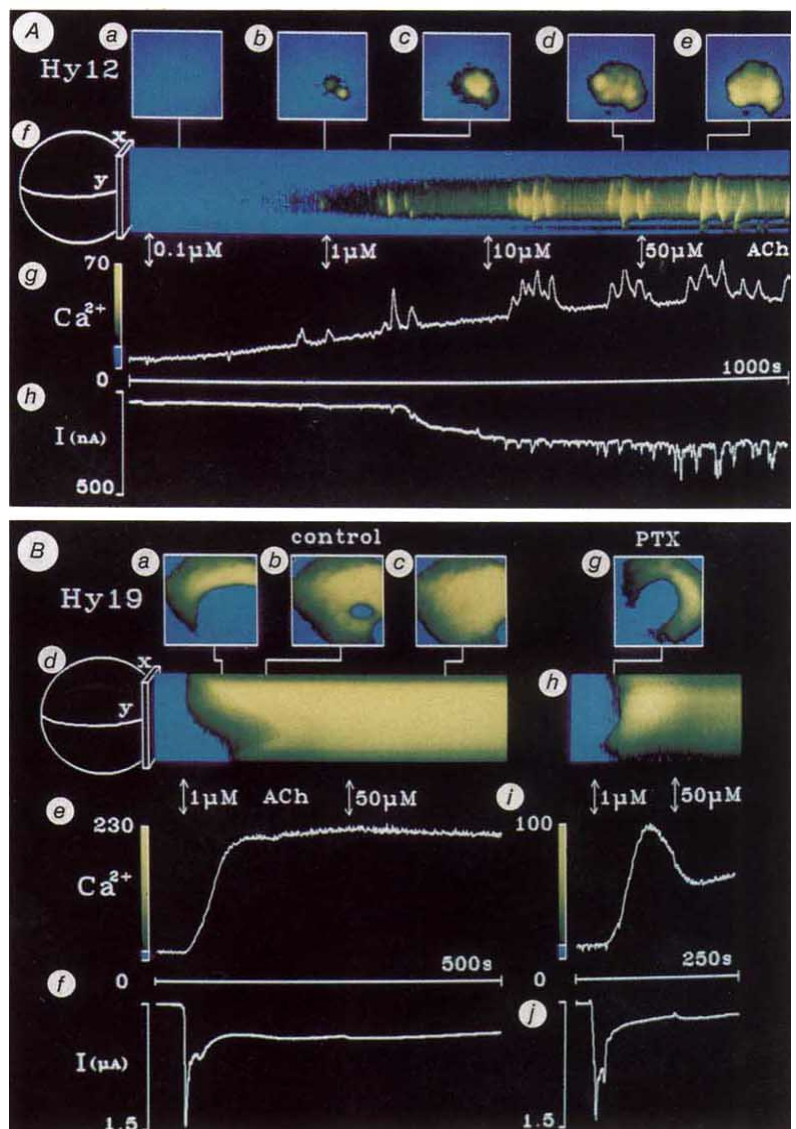


FIG. 3 A, ACh-induced Ca^{2+} release in oocytes expressing Hy12 receptors. *a–e*, Spatial patterns of Ca^{2+} release recorded at 60, 254, 392, 780 and 895 s. Ca^{2+} increases range from 20–70 intensity units. ACh was added at 20 s ($100\ \text{nM}$), 250 s ($1\ \mu\text{M}$), 540 s ($10\ \mu\text{M}$) and at 760 s ($50\ \mu\text{M}$). *f*, Volume rendition of Ca^{2+} release consisting of 1000 sequential images at 1 s intervals. *g, h*, Mean image intensity and $I_{\text{Ca,Cl}}$ during experiment. *B*, ACh-induced Ca^{2+} release in oocytes expressing Hy19 receptors. *a–c*, Spatial patterns of Ca^{2+} release recorded at 106, 163 and 440 s. Ca^{2+} increases range from 20–230 intensity units. ACh was added at 40 s ($1\ \mu\text{M}$) and 280 s ($50\ \mu\text{M}$). *d*, Volume rendition of Ca^{2+} release (500 sequential slices recorded at 1 s intervals). *e, f*, Mean image intensity and $I_{\text{Ca,Cl}}$ during experiment. *g–j*, Ca^{2+} responses in a second oocyte pretreated with PTX ($2\ \mu\text{g ml}^{-1}$, 12 h). Single image recorded at 86 s. Ca^{2+} increases range from 20–100 intensity units. ACh added at 20 s ($1\ \mu\text{M}$) and 150 s ($50\ \mu\text{M}$). *h*, Volume rendition of Ca^{2+} release (250 sequential images). *i, j*, Mean image intensity and $I_{\text{Ca,Cl}}$ during experiment.



Different G-protein pathways mediating each of the characteristic Ca^{2+} release patterns were identified by testing the pertussis toxin (PTX) sensitivity of the m2- and m3-like receptor responses. We found that the characteristic spatial-temporal pattern of Ca^{2+} release mediated by m3 or Hy19 receptors was largely insensitive to PTX treatment ($2 \mu\text{g ml}^{-1}$ for 12 h). Specifically, 14 of 19 oocytes expressing m3 receptors responded with a Ca^{2+} wave when $I_{\text{Ca,Cl}}$ was observed (Table 1). None of these oocytes exhibited any detectable focal activity at saturating ACh concentrations ($50 \mu\text{M}$). More importantly, six of seven oocytes expressing Hy19 receptors produced a large transient $I_{\text{Ca,Cl}}$ and all six of these oocytes exhibited the characteristic m3-like Ca^{2+} waves (Fig. 3B, g-j). The low level of light detection for the m2- and Hy12-coupled Ca^{2+} pathways precluded a meaningful determination of their PTX sensitivity using image analysis. However, we previously demonstrated that $I_{\text{Ca,Cl}}$ evoked by m2 and Hy12 is completely blocked by PTX treatment⁶.

Receptor pathways that are coupled differentially to Ca^{2+} release provide a unique mechanism by which highly convergent receptor signalling can be delineated within cells. Less well coupled pathways will produce more spatially complex patterns of Ca^{2+} release, compared with well-coupled pathways, when challenged with identical agonist concentrations. The lower sensitivity of the m2-mediated Ca^{2+} release pathway could be accounted for by several mechanisms. The m2 and m3 mAChRs could couple to different phospholipase C enzymes¹³, or less efficiently to the same phospholipase C, or they could stimulate an entirely different intracellular Ca^{2+} release pathway. Inhibition of m2 and m3 pathways (>90%; data not shown) by heparin injections (50 international units per oocyte; a nonselective, putative inositol trisphosphate receptor blocker) suggests that both receptors act through an inositol trisphosphate-dependent pathway. Alternatively, the distribution of G proteins or compartmentalized Ca^{2+} stores may determine the ease with which a receptor releases Ca^{2+} .

We conclude that spatial and temporal heterogeneity in Ca^{2+} release patterns define distinct receptor pathways mediated by different G proteins. A propagating Ca^{2+} wave mediated by a PTX-insensitive G protein is induced by m3-like receptor activation, while m2-like receptor activation induces a slow, oscillatory and focal release of intracellular Ca^{2+} mediated by a PTX-sensitive G protein. The prime determinant of the characteristic

Ca^{2+} release patterns for mAChRs cannot be expression levels because we have previously demonstrated that m2 receptors are consistently expressed at levels fourfold higher than m3 receptors in *Xenopus* oocytes⁶ and, more importantly, the PTX sensitivity of the receptor responses is expression-independent. Remarkably, the distinct spatial and temporal patterns of Ca^{2+} release, revealed by hybrid receptor analysis, suggests that receptors generate characteristic profiles of intracellular Ca^{2+} release through the activation of different G proteins. The richly varied subcellular patterns of Ca^{2+} release encoded by different receptor pathways reveal a new level of signalling complexity. □

Received 27 November 1990; accepted 21 January 1991.

1. Berridge, M. J. & Irvine, R. F. *Nature* **341**, 197-205 (1989).
2. Berridge, M. J., Cobbold, P. H. & Cuthbertson, K. S. R. *Phil. Trans. R. Soc. B* **320**, 325-343 (1988).
3. Brooker, G., Seki, T., Croll, D. & Wahlestedt, D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2813-2817 (1990).
4. Rooney, T. A., Sass, E. J. & Thomas, A. P. *J. biol. Chem.* **265**, 10792-10796 (1990).
5. Berridge, M. J. *J. biol. Chem.* **265**, 9583-9586 (1990).
6. Lechleiter, J. et al. *EMBO J.* **9**, 4381-4390 (1990).
7. Fukuda, K. et al. *Nature* **327**, 623-625 (1987).
8. Fukuda, K. et al. *Nature* **335**, 355-358 (1988).
9. Peralta, E. G. et al. *Nature* **334**, 434-437 (1988).
10. Ashkenazi, A. et al. *Cell* **56**, 487-493 (1989).
11. Lechleiter, J., Peralta, E. G. & Clapham, D. *Trends pharmacol. Sci.* (Suppl.) 34-38 (1989).
12. Parker, I. & Ivorra, I. *Proc. natn. Acad. Sci. U.S.A.* **87**, 260-264 (1990).
13. Majerus, P. W. et al. *Cell* **63**, 459-465 (1990).
14. Hernandez-Cruz, A., Sala, F. & Adams, P. R. *Science* **247**, 858-862.
15. Fabiato, A. & Fabiato, F. *J. Physiol., Paris* **75**, 463-505 (1979).

ACKNOWLEDGEMENTS. We thank C. Bliton and the ANALYZE software group (Mayo Foundation) for assistance in imaging analysis, K. Duerson for sharing of unpublished electrophysiology data on Hy19, and P. Camacho for comments on the manuscript. This work was supported by the American Heart Association (J.L.), the NIH (D.C. and E.P.), the Whitaker Foundation (D.C.) and Glaxo Inc. (D.C. and E.P.). E.P. is the recipient of a Searle Scholar Award. D.C. is an Established Investigator of the American Heart Association.

Serine phosphorylation-independent downregulation of cell-surface CD4 by *nef*

J. Victor Garcia & A. Dusty Miller

Program in Molecular Medicine, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, USA

A DECLINE in the T-cell population usually marks the onset of progressive immunological disease in individuals infected with the human immunodeficiency virus (HIV). Because CD4^{+} cells help to coordinate efficient immune responses, some of the defects in the immune function in advanced cases of AIDS may be explained by the disappearance of these cells. Therefore, an understanding of the mechanisms used by HIV to induce the reduction of CD4^{+} cells is important. Here we use a Moloney murine leukaemia virus-based retroviral vector in order to express the *nef* gene of HIV-1 in three lymphocytic cell lines expressing CD4. In all cases we find that cell-surface CD4 expression is inversely related to *nef* expression. However, *nef* does not alter steady-state levels of CD4 RNA or CD4 protein. Also, *nef* can downregulate a CD4 triple mutant (Ser → Ala) that is neither phosphorylated nor downregulated by phorbol esters, indicating that *nef* is acting by a different mechanism.

The human immunodeficiency virus, the aetiological agent of AIDS, selectively infects cells expressing the CD4 molecule¹. The HIV genome is much more complex than that of other retroviruses² and encodes at least seven regulatory proteins. Although our understanding of the function of proteins such as Tat and Rev has advanced greatly, the function of other HIV proteins is not clear³. Nef is an HIV protein of relative molecular mass 27,000 (M_r 27K) encoded by a single open reading frame which overlaps with the 3' long terminal repeat (LTR)⁴. The Nef protein is myristylated and localized in the cytoplasm associated with the intracellular portion of the cell membrane⁵.

TABLE 1 Comparison of $I_{\text{Ca,Cl}}$ responses and calcium release patterns in *Xenopus* oocytes expressing muscarinic receptors

Receptor†	Cl^{-} current* Peak <20s	Pattern of Ca^{2+} release		
		Wave	Foci only	Not detectable
m3	49 of 52	39	4	9
m3 PTX	17 of 19	14	0	5
Hy19	11 of 16	11	2	3
Hy19 PTX	6 of 7	6	0	1
m2	0 of 45	1	28	16
Hy12	0 of 6	1	3	2
Control	0 of 1‡	0	1	0

* Number of oocytes responding electrically to $50 \mu\text{M}$ ACh with the peak current occurring in less than 20 s (see ref. 6).

† Amino-acid sequence of muscarinic receptors in the region controlling G protein selectivity; residues derived from m3 are boxed; *tm5* indicates the proposed cytoplasmic boundary of the fifth transmembrane segment of m2 and m3 receptors.

tm5

m2 VLYWHISRASKSRIKKDKKEPVANQ

m3 ILYWRIYKETEKRTKELAGLQASGT

Hy12 ILYWHISRASKSRIKKDKKEPVANQ

Hy19 VLYWHISRASKSRIKKLAGLQASGT

‡ Control oocytes injected with water, only 1 of 10 responded.

Whether or not Nef has a negative effect on HIV-LTR transcription is arguable⁶⁻⁸. Transient expression of *nef* in a vaccinia virus vector leads to decreased cell-surface expression of CD4 (ref. 9), although others believe that *nef* does not modulate CD4 expression¹⁰.

To examine the effect of *nef* expression on CD4 and its cell-type specificity, a Moloney murine leukaemia virus-based retroviral vector encoding *nef* (*Lnef*SN; Fig. 1a) was constructed by ligating a *Bgl* II fragment of HIV-1 (nucleotide positions 8,504 to 9,573 of the SF2 isolate) into the *Bam*HI site of pLXSN (ref. 11). This fragment contains the entire *nef* coding region as well as 289 base pairs (bp) of 5' and 140 bp of 3' noncoding sequences. Virus made by using PA317 amphotropic retrovirus packaging cells was used to transduce the *nef* gene into human cell lines of T, B and monocyte/macrophage origin (HPBALL, AA2 and U937 respectively), all of which express CD4. Representative populations of G418-resistant cells expressing a 27K protein immunoreactive with rabbit anti-Nef antisera were obtained (Fig. 1b). When analysed for CD4 cell-surface expression by fluorescence-activated cell sorting (FACS), there was a significant downregulation of cell-surface CD4 in all three cell lines expressing *nef*, whereas cells infected with a control

virus showed normal levels of cell surface CD4 (Fig. 1c, e, g). Similar results were obtained with four different monoclonal antibodies directed against CD4, showing that this phenomenon is not due to the masking of a particular epitope on the extracellular portion of the CD4 molecule. In contrast, the same cells transduced with a *tat* or *rev* vector did not show any significant change in CD4 cell-surface levels (data not shown). *Nef* does not downregulate CD8 (Fig. 1h) or the transferrin receptor (data not shown) from the cell surface, indicating that the effect is relatively specific for CD4. To confirm that the HIV-LTR sequences present in *Lnef*SN were not responsible for CD4 cell-surface down regulation, the levels of surface CD4 of cells infected with *Lfen*SN, a retroviral vector with the *Bgl*II fragment containing the *nef* gene inserted in the reverse orientation, were determined. *Lfen*SN-infected cells show no significant differences in surface CD4 levels compared with control cells infected with the LN vector. (Fig. 1i and j).

Whereas U937 cells transduced with *Lnef*SN virus had background levels of cell-surface CD4 (Fig. 1g), the HPBALL and AA2 cells infected with the same virus showed two discrete populations with either low or normal levels of surface CD4 (Fig. 1c, e). To understand this, the two populations (high and

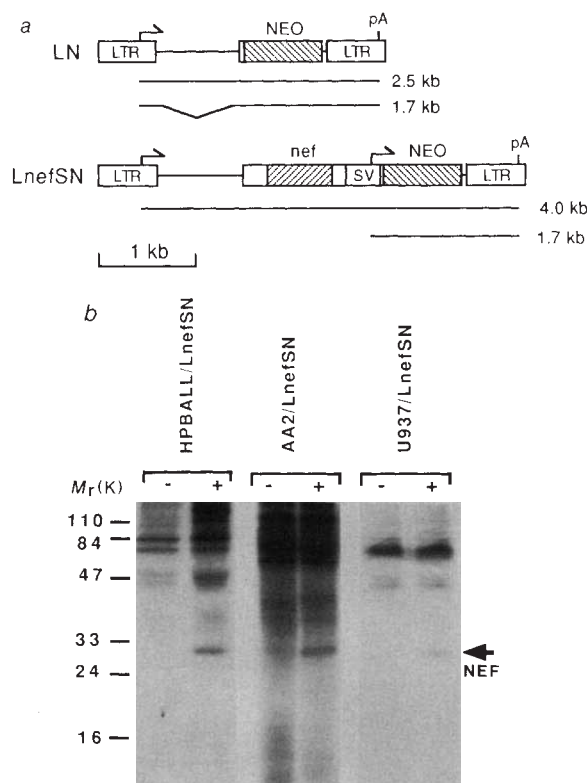
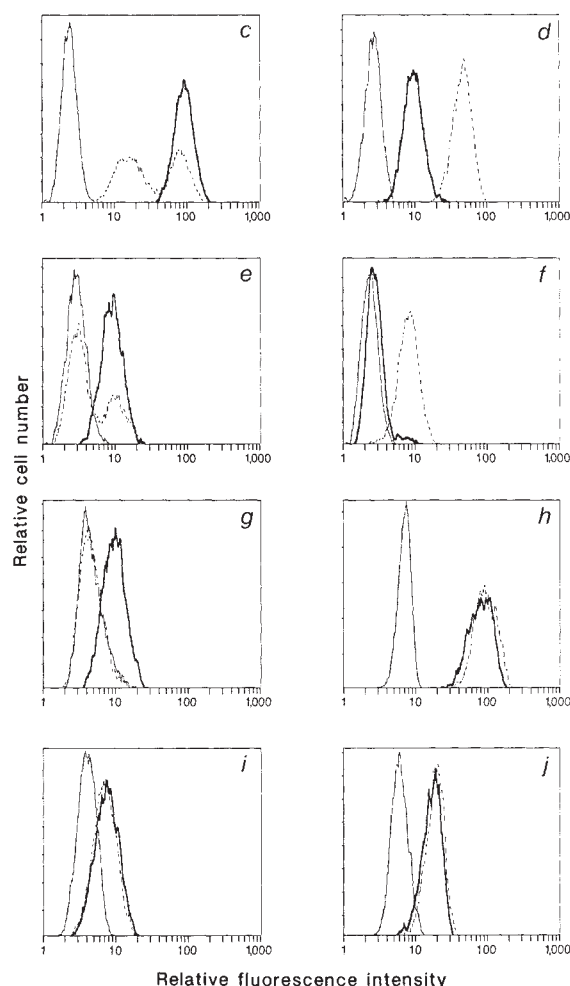


FIG. 1 a, Schematic representation of the LN and *Lnef*SN retroviral vectors. LTR, long terminal repeat; pA, polyadenylation signal; SV, SV40 early promoter; hatched boxes, coding regions. Sizes of the RNA species for LN¹¹ and *Lnef*SN are indicated, assuming 150 residues of poly(A) tail. b, Nef expression in cells infected with *Lnef*SN. Metabolically labelled extracts from cells infected with *Lnef*SN were immunoprecipitated with rabbit anti-Nef antibodies (+) or control serum (-). Prestained size markers are indicated to the left. c-j, FACS analysis of surface membrane CD4 antigen expression in cells transduced with *Lnef*SN. c, HPBALL cells transduced with *Lnef*SN (dotted line) or a control vector, LN (dark solid line). d, HPBALL cells transduced with *Lnef*SN sorted for high (dotted line) and low (dark solid line) CD4. e, AA2 cells transduced with *Lnef*SN (dotted line) or a control vector (dark solid line). f, AA2 cells transduced with *Lnef*SN sorted for high (dotted line) or low (dark solid line) CD4. g, U937 cells transduced with *Lnef*SN (dotted line) or a control vector (dark solid line). h, CD8 expression in HPBALL cells transduced with *Lnef*SN (dotted line) or a control vector (dark solid line). i, AA2 cells transduced with *Lnef*SN (dotted line) or a control vector, LN (dark solid line). j, U937 cells transduced



with *Lfen*SN (dotted line) or a control vector, LN (dark solid line). A macaque-specific anti-CD3 (FN18; c-g and i) or an anti-mouse thy 1.1 (31A; h and j) monoclonal antibody was used as negative control (thin solid line in all inserts).

METHODS. Cells ($1-5 \times 10^5$) were infected with the appropriate virus (0.7-2 ml) and selected in medium containing 2 mg per ml G418 (50% active) in parallel. CD4 expression was monitored with leu3a, OKT4 or LAD.3 and L92.5 monoclonal antibodies. CD8 expression was monitored with monoclonal antibody G10-1. Monoclonal antibodies were FITC-labelled or detected with FITC-conjugated goat anti-mouse IgG.

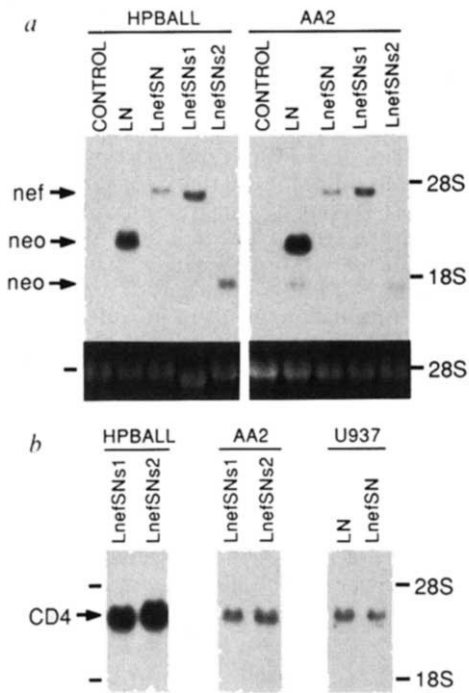
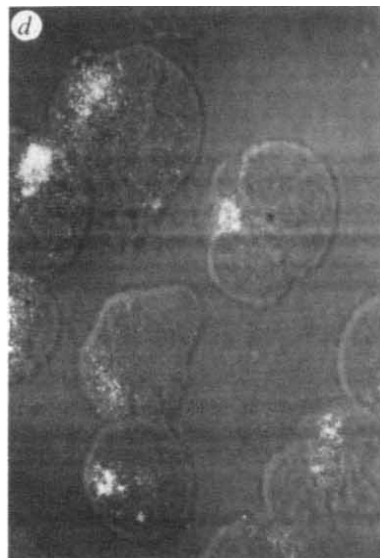
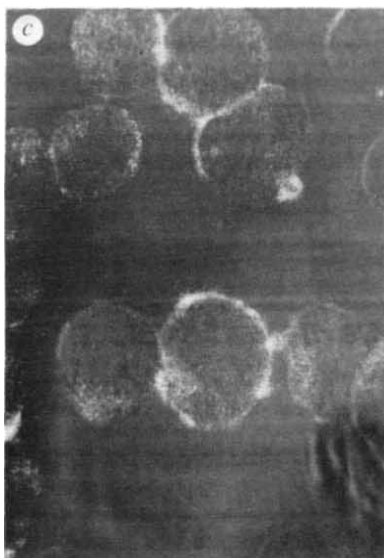
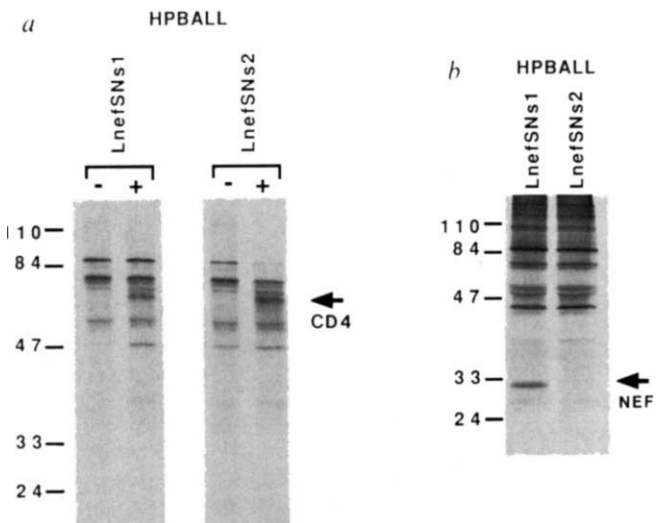


FIG. 2 Analysis of steady-state RNA levels of cells transduced with *Lnef* SN. *a*, Analysis of viral RNAs from HPBALL and AA2 cells uninfected (control), transduced with *Lnef* SN or transduced with a control vector LN¹¹ as well as the sorted populations (S1 and S2, low and high surface CD4 respectively). The positions of the *nef*, *neo*, 28S, and 18S RNAs are indicated (top). Ethidium bromide staining of the 28S RNA (bottom). *b*, Comparison of CD4 RNA levels of cells expressing low (S1) or high (S2) cell surface CD4 levels. The migration positions of the 28S and 18S RNAs are indicated.

METHODS. Total cellular RNA was fractionated on a denaturing agarose gel, stained with ethidium bromide and transferred to a nylon membrane. A *neo*-specific DNA probe was used to identify the viral LTR- (long) or SV40- (short) transcripts (Fig. 1*a*). A CD4-specific DNA probe was used in *b*. HPBALL and AA2 RNA blots in *a* were reprobed with the CD4-specific probe.

FIG. 3 Nef does not affect CD4 biosynthesis. *a*, Total CD4 levels in HPBALL cells expressing low (S1) or high (S2) levels of cell surface CD4. *b*, Nef is only expressed in cells expressing low (S1) levels of cell-surface CD4. In *a* and *b*, prestained size markers are shown to the left (K). *c*, CD4 immunofluorescence staining of HPBALL cells transduced with a control vector (LN). *d*, CD4 immunofluorescence staining of HPBALL cells expressing Nef. *e*, Immunofluorescence staining control with an anti-mouse thy 1.1 monoclonal antibody.

METHODS. Cells (5×10^6) were metabolically labelled with [³⁵S]methionine and [³⁵S]cysteine. *a*, Labelled extracts were immunoprecipitated with anti-CD4 monoclonal antibodies (+) or with a control anti-mouse thy 1.1 monoclonal antibody (-). In *b*, extracts were immunoprecipitated with rabbit anti-Nef antibodies. Samples were fractionated on a 10% (*a*) or a 12.5% (*b*) polyacrylamide gel, dried and exposed to X-ray film. Arrow indicates the position in the gel of the CD4 (*a*) or Nef (*b*) proteins. For immunostaining, cells were fixed with formaldehyde (3.7%) overnight at 4 °C and incubated with normal goat serum. Monoclonal anti-CD4 antibodies (LAD.3 and L92.5) were added at room temperature for 30 min followed by FITC-labelled goat anti-mouse IgG. Images were obtained with an MRC-500 Laser Scanning Confocal Microscope (Bio-Rad Microscience) with a Nikon Optiphot. Phase images were created with the MRC-500 transmission assembly.



low CD4) were separated by FACS. The two populations maintained their phenotype in culture for over four months (Fig. 1d, f). RNA analysis of these two populations showed that the smaller SV40-driven transcript encoding *neo* is present in both populations. However, the LTR-directed transcripts encoding the Nef protein are only present in the cells with low CD4 cell-surface levels (Fig. 2a).

To investigate whether the down-modulation of CD4 was transcriptional, total RNA from cells with high or low levels of surface CD4 was analysed for CD4 messenger RNA. The levels of steady-state CD4 RNA were similar in cells expressing or not expressing the *nef* gene, indicating that *nef* does not alter CD4 RNA levels (Fig. 2b). Therefore it is unlikely that the downregulation of cell-surface CD4 is caused by a negative effect of Nef on CD4 transcription. To test whether Nef influences CD4 biosynthesis translationally, HPBALL/*Lnef*SN cells from the sorted populations were metabolically labelled with medium containing 35 S-labelled methionine and cysteine for 2 h, when serum containing excess unlabelled amino acids was added and incubation continued for 60 min. Surprisingly, both populations synthesize similar amounts of CD4 protein (Fig. 3a), indicating that Nef does not alter CD4 translation. Additionally, the similar migration patterns in both populations suggest that there are no gross differences in CD4 post-translational modification. In both populations longer chase with unlabelled amino acids (up to 5h) did not show significant differences in the rate of CD4 degradation, indicating that Nef does not significantly alter the rate of CD4 degradation (data not shown). The presence of Nef in HPBALL cells expressing low CD4 cell-surface levels and the absence of Nef in HPBALL cells with normal levels of surface CD4 was confirmed by immunoprecipitation of metabolically labelled cell extracts with Nef-specific rabbit antibodies (Fig. 3b).

Indirect immunofluorescence staining was performed on fixed cells using monoclonal antibodies specific for the extracellular domains of CD4. Samples were analysed with a confocal microscope. Fluorescent and phase contrast images were digitized and superimposed to show the cell membrane boundaries. In all cases images represent a 500 nm section through the middle of the cell. Cells not expressing Nef show a pattern of fluorescence typical of cell surface molecules (Fig. 3c). In contrast, cells expressing Nef display almost all fluorescence within the cell (Fig. 3d). The absence of CD4 on the cell surface and its presence inside the cell is consistent with the FACS analysis, which detects only surface CD4 (Fig. 1), as well as with the metabolic labelling experiments (Fig. 3a).

CD4 is downregulated by phorbol esters^{12,13}. Mutation of all intracellular serine residues (at positions 408, 415 and 431) to alanine prevents the phorbol ester-induced downregulation of CD4 (ref. 14). To test whether these serine residues are necessary for Nef-mediated downregulation of CD4, HeLa cells expressing this mutant or wild-type CD4 were transduced with *Lnef*SN. HeLa cells expressing wild-type CD4 infected with a control virus show a normal pattern of CD4 cell-surface expression. However, the same cells transduced with *Lnef*SN show a significant downregulation of their CD4 cell-surface levels (Fig. 4a). A similar result was obtained with cells expressing the CD4 mutant (Fig. 4b). As a control, we confirmed that 12-*O*-tetradecanoylphorbol-13-acetate (TPA) downregulates wild-type CD4, but not the serine triple mutant (data not shown). Therefore, Nef-induced downregulation of cell-surface CD4 expression is independent of these three phosphorylation sites, suggesting that it might occur by a protein kinase C-independent mechanism. Also, because the CD4 genes in these experiments are transcribed by a cytomegalovirus promoter¹⁴, these results suggest that Nef does not act on the CD4 promoter.

Humans infected with HIV develop a degenerative disease of the immune and central nervous systems¹⁵. A decline of the total T-cell population usually marks the onset of progressive immunological disease¹⁶. Our results indicate that HIV is cap-

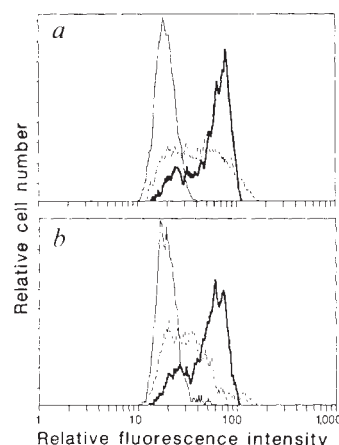


FIG. 4 Downregulation of cell-surface mutant CD4. *a*, HeLa cells expressing wild-type CD4 were transduced with *Lnef* SN (dotted line) or a control virus (dark solid line). *b*, HeLa cells expressing a CD4 triple serine $\rightarrow$ alanine mutant transduced with *Lnef* SN (dotted line) or a control virus (dark solid line). A macaque-specific antibody was used as negative control (thin solid line in *a* and *b*).

METHODS. HeLa cells expressing wild-type CD4 (HeLa.W4) or a triple serine mutant (HeLa.SA1 +2 +3) (ref. 14) were transduced with *Lnef* SN or LN in bulk and selected in G418. After 10 days in selection, colonies of infected cells (>100) were suspended in 1 mM EDTA in PBS, stained with FITC-leu3A, and analysed by FACS.

able of downregulating expression of its own receptor by means of the *nef* gene. Cheng-Mayer *et al.* concluded that *nef* does not downregulate CD4¹⁰. However, the FACS analysis data presented in Fig. 2 of their manuscript shows a significant downregulation of CD4 in their HupAM3 cell line. These data are in agreement with ours and with those reported by Guy *et al.*⁹

Downregulation of cell-surface CD4 by *nef* would be advantageous for the survival and spread of the virus both *in vivo* and *in vitro*. By removing the CD4 molecule from the cell surface, the virus may both effectively impair a normal pathway of T-cell activation and minimize the toxic effects of envelope and superinfection by a second virus^{17,18}. This study demonstrates that HIV Nef protein can downregulate the CD4 molecule from the cell surface of T, B and monocyte/macrophage cell lines as well as non-lymphoid cell lines. Our data show that *nef* expression is sufficient for CD4 cell-surface downregulation by a pathway independent of serine phosphorylation. □

Received 21 December 1990; accepted 11 February 1991.

1. Daigleish, A. G. *et al.* *Nature* **312**, 763–767 (1985).
2. Haseltine, W. A., Terwilliger, D. F., Rosen, C. A. & Sodroski, J. G. in *Retrovirus Biology and Human Disease* (eds Gallo, R. C. & Wong-Staal, F.) (Dekker, New York, 1990).
3. Pavlakis, G. M. & Felberg, B. K. *New Biol.* **2**, 20–31 (1990).
4. Ratner, L. *et al.* *Nature* **313**, 277–284 (1985).
5. Franchini, G., Robert-Guroff, M., Grayeb, J., Chang, N. T. & Wong-Staal, F. *Virology* **155**, 593–599 (1986).
6. Ahmad, N. & Venkatesan, S. *Science* **241**, 1481–1485 (1988).
7. Kim, S., Ikeuchi, K., Byrn, R., Groopman, J. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9544–9548 (1989).
8. Hammes, S., Dixon, E., Mallin, M., Cullen, B. & Greene, W. C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9549–9553 (1989).
9. Guy, B. *et al.* *Nature* **330**, 266–269 (1987).
10. Cheng-Mayer, C., Iannello, P., Shaw, K., Luciw, P. A. & Levy, J. A. *Science* **246**, 1629–1632 (1989).
11. Miller, A. D. & Rosman, G. *Biotechniques* **7**, 980–990 (1989).
12. Hoxie, J. A., Matthews, D. M., Callahan, K. J., Cassel, D. L. & Cooper, R. A. *J. Immunol.* **137**, 1194–1201 (1986).
13. Acres, R. B., Conlon, P. J., Mochizuki, D. Y. & Gallis, B. *J. Biol. Chem.* **261**, 16210–16214 (1986).
14. Shin, J., Doyle, C., Yang, Z., Kappes, D. & Strominger, J. L. *EMBO J.* **9**, 425–434 (1990).
15. Fauci, A. S. *Science* **239**, 617–622 (1988).
16. Gallo, R. C. *J. acq. Immun. Defic. Synd.* **3**, 380–389 (1990).
17. Temin, H. M., Keshet, E. & Weller, S. K. in *Cold Spring Harb. Symp. quant. Biol. Vol. XLIV* 773–778 (Cold Spring Harbor Laboratory, New York, 1980).
18. Weller, S. K., Joy, A. E. & Temin, H. M. *J. Virol.* **33**, 494–506 (1980).

ACKNOWLEDGEMENTS. We thank L. Rose, F. Smith, D. Willerford and S. Rozeman for monoclonal antibodies, E. Clark and J. Strominger for cell lines and J. Alfano for technical assistance. The AA2 (M. Hershfield) and U937 cell lines as well as the rabbit anti-Nef antibodies were obtained from the AIDS repository of the NIAID. We thank H. Weintraub, M. Linial, J. Cooper and M. Emerman for critical reading of this manuscript, J. Torgerson for typing it, and M. Kaleko for his advice throughout this work. This work was supported by grants from the NIAID.

A novel cyclin encoded by a *bcl1*-linked candidate oncogene

Toru Motokura*, Theodora Bloom†, Hyung Goo Kim*, Harald Jüppner*, Joan V. Ruderman†, Henry M. Kronenberg* & Andrew Arnold*‡

*Endocrine Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

†Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

‡To whom correspondence should be addressed

WE have previously identified a candidate oncogene (PRAD1 or D11S287E) on chromosome 11q13 which is clonally rearranged with the parathyroid hormone locus in a subset of benign parathyroid tumours^{1,2}. We now report that a cloned human placental PRAD1 complementary DNA encodes a protein of 295 amino acids with sequence similarities to the cyclins. Cyclins can form a complex with and activate p34^{cdc2} protein kinase, thereby regulating progress through the cell cycle (reviewed in refs 3–5). PRAD1 messenger RNA levels vary dramatically across the cell cycle in HeLa cells. Addition of the PRAD1 protein to interphase clam embryo lysates containing inactive p34^{cdc2} kinase and lacking endogenous cyclins allows it to be isolated using beads bearing p13^{sucr}, a yeast protein that binds cdc2 and related kinases with high affinity and coprecipitates kinase-associated proteins. Addition of PRAD1 also induces phosphorylation of histone H1, a preferred substrate of cdc2. These data suggest that PRAD1 encodes a novel cyclin whose overexpression may play an important part in the development of various tumours with abnormalities in 11q13.

In three independent parathyroid adenomas with similar clonal chromosomal inversions^{6,7} the 5' regulatory region of the parathyroid hormone gene (PTH; on 11p15) is juxtaposed to initially unidentified DNA (D11S287) normally located on 11q13 (refs 1, 2 and Fig. 1a). A D11S287 genomic fragment, probe B, (ref. 1 and Fig. 1b) detects a transcript (D11S287E or PRAD1, for parathyroid adenomatosis) that is greatly overexpressed in these tumours². These clonal rearrangements associated with altered gene expression indicate that PRAD1 may be an oncogene; its juxtaposition near a potent tissue-specific enhancer is analogous to rearrangements of *c-myc* or *bcl2* oncogenes with immunoglobulin gene sequences in B-lymphoid tumours (reviewed in ref. 2).

We cloned a normal PRAD1 cDNA by screening a λ gt11 human placental cDNA library with probe B. Several overlapping clones were obtained (Fig. 1c); the sequence of PRAD1 cDNA is shown in Fig. 2a. Ribonuclease protection and primer extension analyses indicated that the 5' end of the mRNA is 10 nucleotides further upstream (data not shown). The longest open reading frame, starting at the first ATG codon, encodes a protein of 295 amino acids of relative molecular mass 33,729. Screening the Genbank peptide database with this sequence revealed significant homology only to members of the cyclin family, with greatest similarity in the region conserved among cyclins, ranging from 19.1% to 33.6% identity, and 44.1% to 59.2% similarity (Fig. 2b).

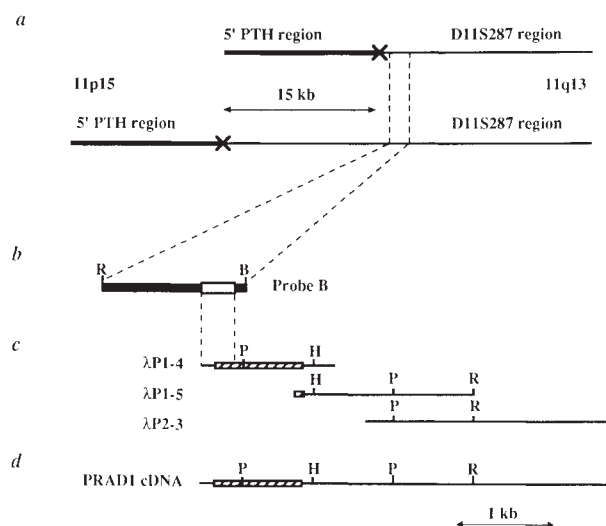
Cyclins were originally identified by their cyclic accumulation and destruction at defined points in embryonic cell cycles⁸. They bind to inactive cdc2 protein kinase and are essential for kinase activation (reviewed in refs 3–5). They can be divided into three families, on the basis of their kinetics of oscillation across the cell cycle, their amino-acid sequences and, in some cases, genetic experiments in yeast that determine when their functions are needed (reviewed in refs 4, 9 and 10). The B-type 'mitotic' cyclins drive cells into mitosis and are conserved from yeast to human^{4,10,11}. The A-type cyclins may act earlier in the cell cycle^{12–14}. The recently described CLNs (or 'G1 cyclins') of budding yeast are believed to have analogous functions by interacting with cdc2 homologues at Start, driving cells into S phase⁹. A, B and CLN cyclins may act as stage-specific regulators of progress across the cell cycle by conferring selective substrate specificity on cdc2 kinase¹² or by selectively targeting cdc2 to different intracellular compartments. The PRAD1 protein has significant sequence similarities to all three types of cyclins, but cannot readily be assigned to any one type. This suggests that PRAD1 represents a new and different cyclin family.

PRAD1 mRNA is expressed in many tissues and is highly conserved, at high stringency the human probe detects transcripts in bovine and murine tissue (Fig. 3a). Some human tissues contain additional, smaller transcripts which may arise by alternative splicing (C. Rosenberg *et al.*, manuscript submitted). As with other cyclin mRNAs expressed in human cells^{11,13}, PRAD1 mRNA levels vary across the cell cycle (Fig. 3b,c), suggesting a role for PRAD1 in cell cycle regulation. The peak in PRAD1 mRNA levels occurs in G2/M or in G1; attempts to monitor protein levels across the cell cycle using an anti-PRAD1 antiserum have not yet been successful.

We have also begun to examine the behaviour of the PRAD1 protein. Cyclins form complexes with the p34^{cdc2} protein kinase

FIG. 1 a, Schematic representation of PTH-D11S287 rearrangements in two parathyroid adenomas^{1,2}. The 5' PTH region (11p15, thick lines) was juxtaposed to the D11S287 region (11q13, thin lines). The breakpoints in the D11S287 region are 15 kb apart. b, Genomic probe B (ref. 1) is shown as a darkened box, whose open area represents the first exon of PRAD1. c, Restriction maps of the inserts of three representative overlapping PRAD1 cDNA clones, λ P1-4, λ P1-5, and λ P2-3. d, Deduced restriction map of the PRAD1 cDNA. The coding region is shown as a crosshatched box. Scale of 1 kilobase is shown as arrows. Symbols used for restriction sites are: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pst*I.

METHODS. b and c, A λ gt11 placental cDNA library (Clontech) was screened with radiolabelled probe B. λ P1-4 and another similar phage clone were isolated. Probe B and the insert of λ P1-4 were sequenced. The region of genomic and cDNA overlap was followed in probe B by a GT splice donor sequence in only one orientation, confirming prior hybridization data² suggesting transcription in the left to right orientation as shown. The next probe was made by polymerase chain reaction amplification of the 3' region of the λ P1-4 cDNA insert and used to rescreen the same library. From 5×10^5 plaque forming units (PFU) of this library, one of 16 positive clones, λ P1-5, had an insert extending further downstream. The *Pst*I-*Eco*RI fragment of λ P1-5 was then used to rescreen the library and 12 similar clones, the longest of which was λ P2-3, were obtained. The sequence of the insert of λ P2-3 revealed polyadenylation signals and a poly(A) stretch of 16 nucleotides in an appropriate position, consistent with the expected orientation. Standard methods for library screening and probe labelling were used²⁴.



[illegible]***b***

human cyclin A: MRAILVDWLVVEVGEEYKIQNECHLAVNYIDRFLSSMSVLRGLQLVGTAAMLASKFEELIYPPEVAEFVYITDDTYTK 268
 * * * * *
pradi: MKRIVATWMLLEVECEQCEEEVFP LAMNYIDRFLSLEPVKRSRLQLGATCMFVASKMKETIPLTAEKLCIYDTSIRP 134
 * * * * *
 clam cyclin A: MRCILDWLVLEVEEDKLHRETLPGLVNYIDRFLSKISVLRGLQLVGAAASMF LAKEYEIPDPVKEFYITDDTYS 273
 * * * * *

[illegible][illegible][illegible]

prad1: MKKIVATMMLEVCCEEQKCEEVFLPAMNYLDRFLSLLEPVKKSRLLQLLGATCMFVASKMKETIPLTA....EKLCIYTD 129
|| ** *** |* ||** |* |||* |****| ** |* || *

cln3: MRFLIFDFIMYCHTRNLNSTSTLFTLTFTILDKYSSRFIIKSYNYQLLSLTALWISSKFWDKSNRMATLKVQLNLC.CNQ 184

prad1: NSIRPEELLQELLVNLKLNLAAMTPHD.FIEHFLSKMPEAEENKQIIRKHAQTFVALCATDVKFISNPPSMAAAGS 207
||* * ||* | || * * ||* || * * * * / * * / * * *
cIn3: YSIK..QFTTMMHFLKSLDWSICQSAFTDSYIDIFLQSTSPSPGVVL..SAPLEAFIQOKLALLNNAAGTAINKS 258

Fig. 2. *a*, Nucleotide sequence and predicted amino-acid sequence (single-letter amino-acid code) of PRAD1 cDNA. The initiation codon, AUUUA motifs (consistent with the rapid change of its mRNA levels) and polyadenylation signal are underlined. Nucleotide numbers are on the left and amino-acid numbers are on the right. Nucleotide 3,495, shown as W, indicates A or T because the sequences of two independent clones did not agree. Nucleotide 4,017 is shown as R, meaning A or G, for the same reason. *b*, Sequence homology in the 'cyclin box' region between the predicted PRAD1 protein and A-type cyclins (human and clam cyclin A, 32.2% and 33.6% identity, 56.6% and 59.2% similarity)^{14,22}, B-type cyclins (human cyclin B and *Saccharomyces pombe cdc13*, 29.6% and 31.6% identity, 50.0% and 52.0% similarity)^{11,25}, and one *S. cerevisiae* G1 cyclin (*cln3*, 19.1% identity, 44.1% similarity)^{26,27}. Clam cyclin A and *S. pombe cdc13* homologies with PRAD1

are representative of those found in their families; *clin3* is better aligned with PRAD1 than is *clin1* or 2. Identical amino acids, |; Conservative substitutions, *. Alignment was made with the assistance of the BESTFIT program²⁸ and conservative amino acids are grouped as follows: D, E, N, Q; H, K, R, A; G, P, S, T; I, L, M, V; F, W, Y. Amino acid numbers are on the right.

METHODS. The inserts of the clones, λ P1-4, λ P1-5 and λ P2-3 shown in Fig. 1c, and of other independent clones, were subcloned into pGEM7ZF(+) (Promega). Sequence was obtained using the double-stranded DNA sequencing technique (dideoxy chain terminator method) with Sequenase (USB) as described by the manufacturer. Several oligonucleotides were synthesized as internal primers to facilitate sequencing. The coding region was sequenced in both orientations and in at least two independent clones.

leading to its activation, which can be assayed using exogenous histone H1 as a substrate. In addition, cyclin-p34^{cdc2} complexes can be purified using beads linked to p13^{suc1}, another cell cycle protein, which bind p34^{cdc2} and related kinases with high affinity and copurify any proteins complexed with them (ref. 5). When bacterially expressed PRAD1 was added to clam embryo interphase cell lysates¹⁵ (which lack endogenous cyclins and contain inactive p34^{cdc2}), both p34^{cdc2} and PRAD1 were bound by p13^{suc1}-beads (Fig. 4a-c). As PRAD1 is not bound to protein A/Sepharose beads, its binding to p13^{suc1}-beads is probably due to its interaction with p34^{cdc2} or a closely related protein. Furthermore, kinase activity was induced by the addition of the PRAD1 *in vitro* translation product to interphase lysates (Fig. 4d). This kinase activity was less than that seen with cyclin A. Cyclin B provided a negative control; for reasons not yet understood, our cyclin B translation product was not capable of inducing kinase activity in this type of assay (unpublished observation). The difference between the activities induced by cyclin A and PRAD1 may be specific to this clam assay system, or may reflect a genuine difference between the functions of, or the substrate specificities conferred by, cyclin A and PRAD1. Because of its sequence and these preliminary observations of

its behaviour, we propose that PRAD1 is a novel, cyclin-related regulator of the cell cycle.

PRAD1 has been implicated in non-parathyroid neoplasia. An 11q13 region that usually includes *int2*, *hst1* and *bcl1* is amplified in 15–20% of human breast and squamous cell carcinomas^{16–19}. PRAD1 is invariably amplified and, unlike the other loci, also overexpressed in these tumours²⁰. In B-cell neoplasia²¹ *bcl1* was identified as the 11q13 breakpoint region in the chromosomal translocation t(11;14)(q13;q32)²¹; the existence of a *bcl1* region oncogene has been assumed but is not yet identified. In view of recent findings that PRAD1 is ≤ 240 kilobases (kb) from *bcl1* (ref. 20 and A. Bale, personal communication) and is overexpressed in B-cell tumours characterized by *bcl1* rearrangement (C. Rosenberg *et al.*, manuscript submitted), PRAD1 may be the *bcl1* oncogene.

Deregulated expression of other cyclins may also be involved in abnormal cell proliferation. In a hepatocellular carcinoma, the human cyclin A gene was identified as the site of clonal integration of hepatitis B virus²². In adenovirus-transformed cells, human cyclin A is associated with the transforming viral protein, E1A (ref. 13). Finally, potential substrates of p34^{cdc2} kinase include p60^{c-src}, the large T antigen of simian virus 40,

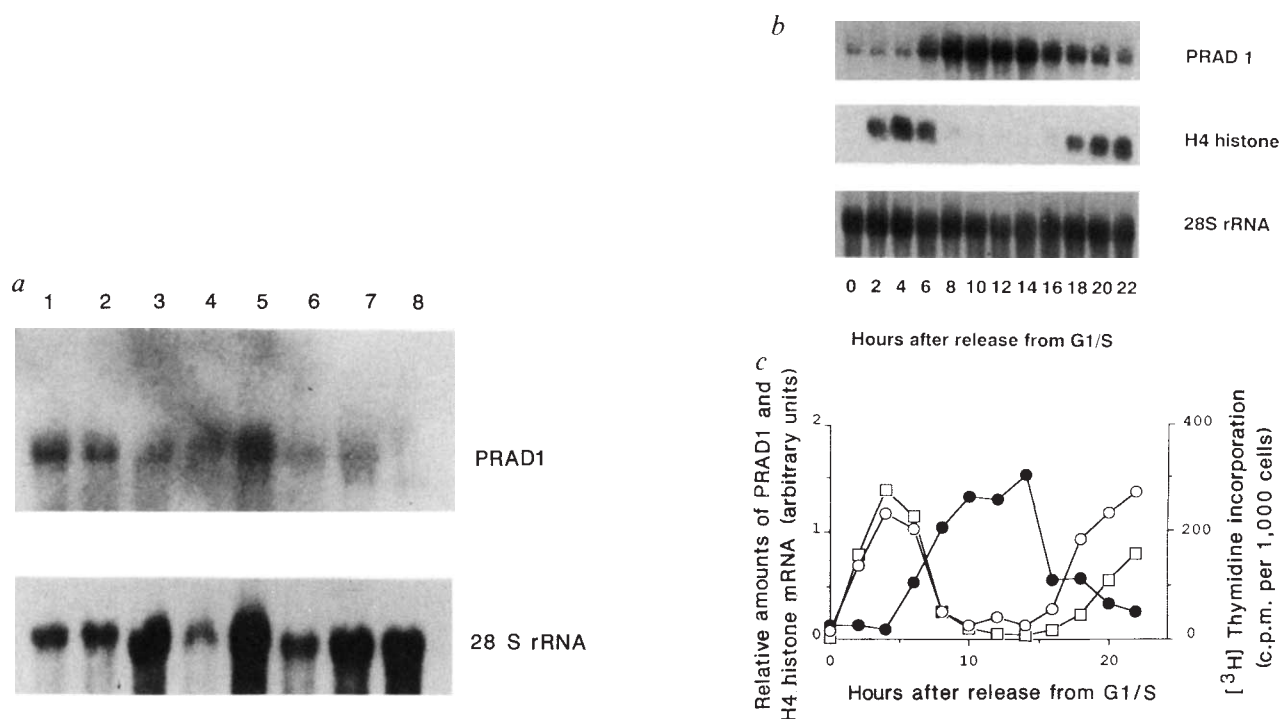


FIG. 3 a, Northern blot analysis of total RNA from human thyroid (lane 1), human placenta (lane 2), bovine parathyroid (lane 3), bovine thyroid (lane 4), bovine lymph node (lane 5), bovine skeletal muscle (lane 6), murine heart (lane 7) and murine liver (lane 8). PRAD1 mRNA shown in the upper panel is of ~ 4.5 kilobases, slightly smaller than the 28S ribosomal rRNA; the mouse PRAD1 mRNA is smaller than the human or bovine mRNAs. 28S rRNA hybridization is shown in the lower panel. b, Northern blot analysis of total RNA from HeLa S3 cells after release from G1/S block. PRAD1 mRNA is shown in the upper panel. H4 histone mRNA in the middle panel shows the pattern expected in well synchronized cells²⁹ and 28S rRNA is shown in the lower panel as a control for RNA loading. c, Relative amounts of PRAD1 mRNA (—●—), H4 histone mRNA (—○—) and [³H]thymidine incorporation (—□—) of HeLa S3 cells after release from G1/S block. The signals of the blot shown in a were measured by densitometry and normalized to the 28S rRNA.

METHODS. a, RNAs were prepared from the indicated tissues by standard procedures²⁴ and 10 μ g total RNA was loaded, separated on an agarose-formaldehyde gel, blotted onto nitrocellulose, hybridized with probe B and the 28S rRNA oligonucleotide². The filter was washed at high stringency

(0.1 $\times$ SSC, 60 $^{\circ}$ C) and autoradiographed. b, c, HeLa S3 cells (American Type Culture Collection), maintained in DMEM medium (GIBCO) with 7% fetal bovine serum (FBS), were synchronized at the G1/S boundary by sequential thymidine-aphidicolin treatment²⁹ with a slight modification. Log-phase cells were incubated in complete medium (DMEM with 7% FBS, penicillin G and streptomycin) with addition of thymidine (Sigma) 2 mM for 12 h. After release from thymidine block by three washes with PBS, the cells were incubated for 10 h with deoxycytidine (Sigma, 24 μ M) and thymidine (24 μ M), trypsinized, counted and aliquoted equally (5×10^4 cells cm^{-2}). Incubation with aphidicolin (Sigma; 5 μ g ml^{-1} for 14 h) was followed by release from G1/S block with four washes with DMEM and incubation in complete medium. [³H]thymidine (NEN) was added to an aliquot 15 min before each indicated time point; a 30-min incubation and collecting for TCA precipitation followed. RNAs from parallel aliquots were extracted³⁰ at the indicated times; 0 h was just before release from aphidicolin. RNAs (5 μ g per lane) were blotted onto nitrocellulose and sequentially hybridized with the PRAD1 λ p1-4 cDNA insert, human H4 histone pF0108X (ref. 31) and a 28S rRNA oligonucleotide as described above.

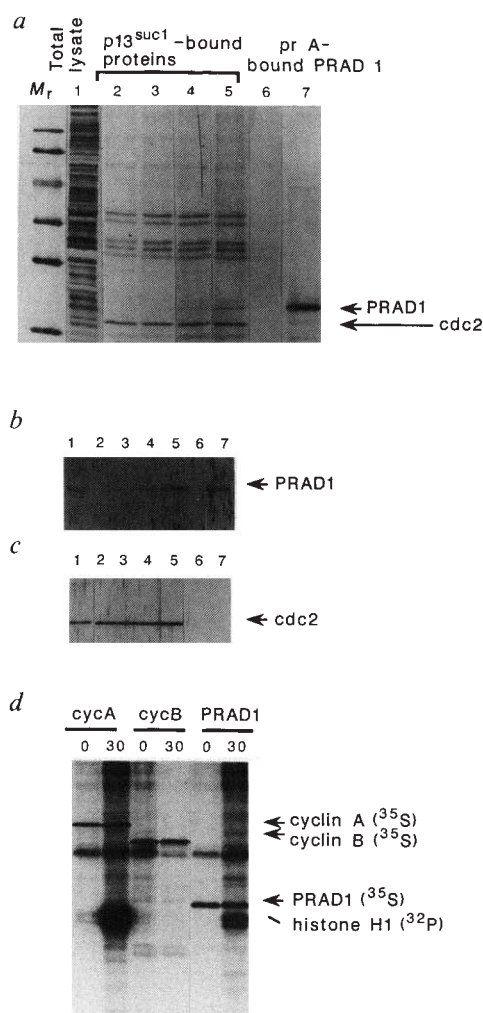


FIG. 4 PRAD1 protein added to clam embryo cell lysates binds to $p13^{suc1}$ -Sepharose beads and activates histone H1 kinase activity. *a, b, c*, Bacterially expressed PRAD1 was incubated with a clam embryo interphase lysate lacking endogenous cyclins A and B. The lysates were then mixed with $p13^{suc1}$ -Sepharose or protein A-Sepharose beads. The bound material was eluted, electrophoresed and either silver stained (*a*) or immunoblotted with anti-PRAD1 antiserum (*b*) or anti-*cdc2* antiserum (*c*). Lane M shows relative molecular mass (M_r) markers (from top to bottom) of 116K, 94K, 68K, 56K, 40K and 31K. Lane 1 shows whole clam embryo interphase lysate plus 18 ng PRAD1 protein. Lanes 2, 3, 4, 5 and 6 represent clam embryo lysate to which 0, 18, 45, 225 or 18 ng PRAD1, respectively, were added; these mixes were assayed for material binding to $p13^{suc1}$ -Sepharose (lanes

2–5) or protein A-Sepharose (lane 6) beads. Lane 7 shows bacterially expressed PRAD1. Arrows indicate the positions of PRAD1 and *cdc2* marker proteins. The identities of the larger proteins also bound by $p13^{suc1}$ beads are not known. *d*, Equal volumes of clam embryo interphase lysate and reticulocyte lysate containing ^{35}S -labelled clam cyclin A (*cycA*), cyclin B (*cycB*) or human PRAD1 were mixed and incubated at 18 °C. Samples were taken at 0 and 30 min and assayed for histone H1 kinase activity. The final reaction products, which include both ^{35}S -labelled proteins and ^{32}P -labelled kinase products, were then examined by SDS-PAGE followed by autoradiography.

METHODS. Clam embryo interphase cell lysates lacking endogenous cyclins were prepared by adding 100 μM emetine during first mitosis, as described previously¹⁵, followed by homogenization and centrifugation at 150,000g. Aliquots of the supernatant were frozen in liquid nitrogen. [^{35}S]methionine-labelled PRAD1 was produced in reticulocyte lysate *in vitro* translation system (Promega) according to manufacturer's instructions by using the plasmid, pP1-8, containing the $\lambda\text{P1-4}$ insert in pGEM7Zf(+) (Promega). To produce PRAD1 in *Escherichia coli*, pT4R-1 was constructed by insertion of the $\lambda\text{P1-4}$ insert into the *NcoI* and *BamHI* sites of pET-3d (ref. 32). BL21(DE3) cells were transformed with pT4R-1, cultured and treated with 0.4 mM IPTG for 3 h to induce PRAD1 expression. The induced product was purified from cell culture as inclusion bodies, solubilized in 9M urea and dialysed against PBS (ref. 35). By SDS-PAGE the apparent sizes of the *in vitro* translation product and the bacterially expressed product were the same (M_r 35K) (data not shown). Rabbit anti-PRAD1 antisera were raised against a synthetic peptide corresponding to amino acids 9–37 of PRAD1. Antisera were assayed by immunoprecipitation of the *in vitro* translation product. Antisera specificity was shown by comparison with normal rabbit serum and by successful competition with the (9–37) peptide (data not shown). *a, b, c*, Thawed clam embryo lysate (16.5 μl) and bacterially expressed PRAD1 (5.5 μl) were mixed and incubated at 18 °C for 30 min before transfer to 4 °C, dilution with four volumes of buffer A (50 mM Tris, pH 7.4, 150 mM NaCl, 5 mM EDTA, 5 mM EGTA, 1 mM ammonium molybdate) and addition of $p13^{suc1}$ -Sepharose or protein A-Sepharose followed by mixing for 1 h. Beads were then pelleted, washed in buffer A + 0.5% Tween-20, in buffer B (50 mM Tris, pH 7.4, 1 M NaCl, 5 mM EDTA, 5 mM EGTA, 1 mM molybdate, 0.5% Tween-20) and finally in buffer A without Tween-20, all at 4 °C. Washed beads were boiled in SDS sample buffer for 3 min and the supernatant split into three samples for electrophoresis. Gels were silver-stained or blotted onto nitrocellulose filters and reacted with rabbit antibodies generated against bacterially expressed full-length *S. pombe cdc2* protein (E. Shibuya and J.V.R., manuscript in preparation) or PRAD1 peptide as above. Antibody binding was visualized by alkaline phosphatase-linked secondary antibodies according to the manufacturer's (Promega) directions. *d*, Synthetic clam cyclins A and B (refs 10, 14) and PRAD1 mRNAs were transcribed and translated as described above. Translation product (3 μl) and clam embryo lysate (3 μl) were mixed. Samples were frozen immediately in liquid nitrogen. The remainder was incubated for 30 min at 18 °C and then frozen. Samples were diluted with 1 volume of ice-cold buffer A, thawed on ice and mixed with an equal volume of kinase mix (40 mM HEPES, pH 7.3, 20 mM MgCl_2 , 10 mM EGTA, 0.2 mg per ml histone H1, 10 μM cAMP-dependent kinase inhibitor (Sigma), 0.5 mCi per ml [$\gamma\text{-}^{32}\text{P}$]ATP and incubated at 23 °C for 10 min. Double-strength SDS sample buffer was then added and the entire mix was analysed by electrophoresis followed by staining with Coomassie blue and autoradiography. The position of the stained histone H1 protein is marked.

c-abl and the tumour suppressor gene product p53 (ref. 23). Knowledge of these links between cell cycle control and oncogenesis may be considerably strengthened by further investigation of PRAD1. □

Received 4 March; accepted 22 March 1991.

- Arnold, A. *et al.* *J. clin. Invest.* **83**, 2034–2040 (1989).
- Rosenberg, C. L., Kim, H. G., Shows, T. B., Kronenberg, H. M. & Arnold, A. *Oncogene* **6**, 449–454 (1991).
- Murray, A. W. & Kirschner, M. W. *Science* **246**, 614–621 (1989).
- Nurse, P. *Nature* **344**, 503–508 (1990).
- Draetta, G. *et al.* *Cell* **56**, 829–838 (1989).
- Arnold, A., Staunton, C. E., Kim, H. G., Gaz, R. D. & Kronenberg, H. M. *New Engl. J. Med.* **318**, 658–662 (1988).
- Friedman, E. *et al.* *J. clin. Endocr. Metab.* **71**, 293–297 (1990).
- Evans, T., Rosenthal, E. T., Youngblood, J., Distel, D. & Hunt, T. *Cell* **33**, 389–396 (1983).
- Nasmyth, K. A. *Cell* **63**, 1117–1120 (1990).
- Westendorp, J. M., Swenson, K. I. & Ruderman, J. V. *J. Cell Biol.* **108**, 1431–1444 (1989).
- Pines, J. & Hunter, T. *Cell* **58**, 833–846 (1989).
- Minshall, J., Golsteyn, R., Hill, C. & Hunt, T. *EMBO J.* **9**, 2865–2875 (1990).
- Pines, J. & Hunter, T. *Nature* **346**, 760–763 (1990).
- Swenson, K. I., Farrell, K. M. & Ruderman, J. V. *Cell* **47**, 861–870 (1986).

- Luca, F. C. & Ruderman, J. V. *J. Cell Biol.* **109**, 1895–1909 (1989).
- Ali, I. U., Merlo, G., Callahan, R. & Lidereau, R. *Oncogene* **4**, 89–92 (1989).
- Fantl, V. *et al.* *Eur. J. Cancer* **26**, 423–429 (1990).
- Theillet, C. *et al.* *Oncogene* **5**, 147–149 (1990).
- Berenson, J. R., Yang, J. & Mickel, R. A. *Oncogene* **4**, 1111–1116 (1989).
- Lammie, G. A. *et al.* *Oncogene* **6**, 439–444 (1991).
- Tsujiyama, Y. *et al.* *Science* **224**, 1403–1406 (1984).
- Wang, J., Chenivess, X., Henglein, B. & Bréchet, C. *Nature* **343**, 555–557 (1990).
- Hunter, T. *Cell* **64**, 249–270 (1991).
- Davis, L. G., Dibner, M. D. & Battey, J. F. *Basic Methods in Molecular Biology* (Elsevier, New York, Amsterdam, London, 1986).
- Bohrer, R. & Beach, D. *EMBO J.* **7**, 2321–2327 (1988).
- Nash, R., Tokiwa, G., Anand, S., Erickson, K. & Fletcher, A. B. *EMBO J.* **7**, 4335–4346 (1988).
- Cross, F. R. *Molec. cell. Biol.* **8**, 4675–4684 (1988).
- Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
- Heintz, N., Sive, H. L. & Roeder, R. G. *Molec. cell. Biol.* **3**, 539–550 (1983).
- Chomczynski, P. & Sacchi, N. *Analyst. Biochem.* **162**, 156–159 (1987).
- Pauli, U., Chrysogelos, S., Stein, G., Stein, J. & Nick, H. *Science* **236**, 1308–1311 (1987).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **108**, 60–89 (1990).
- Gardella, T. J. *et al.* *J. Biol. Chem.* **265**, 15854–15859 (1990).

ACKNOWLEDGEMENTS. We thank G. and J. Stein for the H4 histone plasmid, pF0108X, H. Keutmann for preparing PRAD1 synthetic peptides and F. Luca, C. Dohrmann, X. Zheng, K. Gavin and E. Shibuya for reagents and discussion. This work was supported by the NIH, the American Cancer Society and the Leukemia Research Foundation.

Activation of the $\beta 1$ isozyme of phospholipase C by α subunits of the G_q class of G proteins

Stephen J. Taylor^{*†}, Ho Zoon Chae[‡], Sue Goo Rhee[‡], & John H. Exton^{*§}

^{*}Howard Hughes Medical Institute, Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

[‡]Laboratory of Biochemistry, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

MANY hormones, neurotransmitters and growth factors, on binding to G protein-coupled receptors or receptors possessing tyrosine kinase activity, increase intracellular levels of the second messengers inositol 1,4,5-trisphosphate and 1,2-diacylglycerol. This is due to activation of phosphoinositide-specific phospholipase(s) C (PLC), the isozymes of which are classified into groups, α , β , γ and δ (refs 1, 2). The β , γ and δ groups themselves contain PLC isozymes which have both common and unique structural domains³. Only the $\gamma 1$ isozyme has been implicated in a signal transduction mechanism⁴. This involves association with, and tyrosine phosphorylation by, the ligand-bound epidermal growth factor and platelet-derived growth factor receptors⁵⁻⁷, probably by means of the PLC- $\gamma 1$ -specific *src* homology (SH2) domain⁸. Because EGF receptor-mediated tyrosine phosphorylation of PLC- $\gamma 1$ stimulates catalytic activity *in vitro*⁹ and G proteins have been implicated in the activation of PLC¹⁰, we investigated which PLC isozymes are subject to G protein regulation. We have purified¹¹ an activated G protein α subunit that stimulates partially purified phospholipase C and now report that this G protein specifically activates the $\beta 1$ isozyme, but not the $\gamma 1$ and $\delta 1$ isozymes of phospholipase C. We also show that this protein is related to the G_q class of G protein α subunits¹².

We reconstituted a purified G protein PLC activator of relative molecular mass 42,000 (M_r 42K; termed GPA-42) from bovine liver membranes and the $\beta 1$, $\gamma 1$ and $\delta 1$ isozymes of PLC purified from bovine brain cytosol. Reconstitution of GPA-42 with PLC- $\beta 1$ markedly stimulated phosphatidyl inositol 4,5-bisphosphate hydrolysis (Fig. 1a). Half-maximal and maximal activation were produced with a molar ratio of GPA-42:PLC- $\beta 1$ of 2:1 and 20:1 respectively. Reconstitution of GPA-42 with either PLC- $\gamma 1$ or PLC- $\delta 1$, however, did not change the enzyme activity (Fig. 1b and c), even at molar ratios $>20:1$. Although the $\beta 1$ and $\delta 1$ isozymes were assayed at a free Ca^{2+} concentration of 0.2 μM (see Fig. 2), PLC- $\gamma 1$ activity was measured at 1 μM Ca^{2+} because of its lower specific activity. However, there was no activation of this isozyme at 0.2 μM Ca^{2+} (data not shown).

Other studies have suggested that G protein-mediated stimulation of PLC activity results in a decreased Ca^{2+} requirement for enzyme activity. PLC- $\beta 1$ was reconstituted with GPA-42 over a range of free Ca^{2+} concentrations (0.02–20 μM), but at all concentrations the stimulation of PLC- $\beta 1$ activity (2.7–3.3 times) was similar (Fig. 2). Maximal stimulation occurred at 1 μM Ca^{2+} in the presence and absence of GPA-42. Even at 0.5 mM Ca^{2+} , activation by GPA-42 was undiminished (data not shown). Therefore G-protein stimulation of PLC- $\beta 1$ activity is probably not a result of an intrinsic increase in affinity for Ca^{2+} . The local membrane environment may, however, influence the interaction of PLC with divalent cations.

To confirm that PLC- $\beta 1$ is a target of GPA-42, bovine liver plasma membranes were incubated with antibodies against the various PLC isozymes. Pretreatment with a mixture of mono-

clonal antibodies or with a polyclonal antiserum specific for the $\beta 1$ isozyme inhibited stimulation of membrane-associated PLC activity by GPA-42 by 84% and 79% respectively, compared with incubation with mouse IgG or control serum (Table 1). Monoclonal antibodies against $\gamma 1$ or $\delta 1$ isozymes gave no inhibition. Anti-PLC- $\beta 1$ serum also decreased stimulation of PLC activity by 10 μM GTP- γ -S (141% of control, compared with 259% of control with control serum $n = 3$; data not shown). Partial inhibition of muscarinic agonist and GTP- γ -S stimulation of rabbit brain membrane PLC activity has also been reported¹³ with antiserum raised against a β -like PLC from the same source.

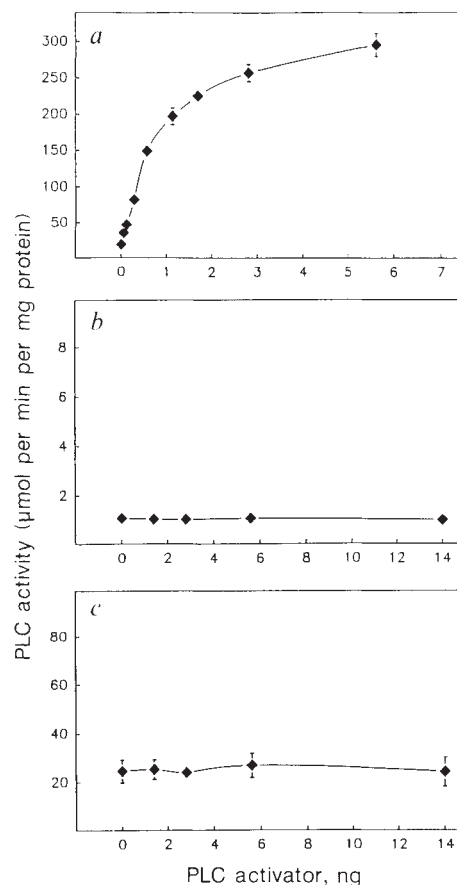


FIG. 1 The effect of GPA-42 on purified PLC isozyme activities. The indicated amounts of GPA-42 (1 ng corresponds to 120 pM) were tested for their ability to modulate PLC activity. a, 1 ng (33 pM) of PLC- $\beta 1$ ($n = 3$); b, 5 ng (169 pM) of PLC- $\gamma 1$ ($n = 3$); c, 1 ng (59 pM) of PLC- $\delta 1$ ($n = 3$).

METHODS. The 42K G protein PLC activator (GPA-42) of bovine liver plasma membranes was purified as described¹¹, with the following modifications. Activation of membranes with GTP- γ -S was performed in the presence of 3 μM 8-arginine vasopressin and 300 μM 5'-adenylylimidodiphosphate and the incubation times was extended to 30 min. Octyl Sepharose chromatography was done on a 60-ml (1.6 $\times$ 30 cm) column and the gradient volume increased to 600 ml; 6-ml fractions were collected. Mono Q chromatographies were performed at pH 7.25; following washing with 80 mM NaCl in buffer C¹¹, the columns were developed with a gradient of 80–316 mM NaCl in buffer C. GPA-42 was pooled on the basis of activity and purity on SDS-PAGE and quantitated by densitometric scanning of a silver-stained gel using purified G protein $\beta \gamma$ subunits as standards. PLC isozymes were purified from bovine brain cytosol as described^{26,27}. Appropriate amounts of GPA-42 or buffer were briefly premixed with PLC isozymes at 4 $^{\circ}C$ prior to assay. For PLC assays¹¹, 0.1 mM [³H]phosphatidyl inositol 4,5-bisphosphate was used as substrate in the presence of a fourfold molar excess of phosphatidylethanolamine. The final concentration of detergent in assays was kept constant and was 0.04% (b, c) or 0.06% (a) cholate, and 0.02% (a) or 0.05% (b, c) octylglucoside. Free Ca^{2+} concentrations were maintained at 0.2 μM (a, c) or 1 μM (b) with EGTA buffers, and assays were for 8 min (a) or 20 min (b, c).

[†] Present address: Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853, USA.

[§] To whom correspondence should be addressed.

TABLE 1 Effect of anti-PLC antibodies on PLC activation in plasma membranes

Preincubation	Addition to assay	
	None	GPA-42
(a) mouse 1gG	2.99 ± 0.47	12.77 ± 4.24
anti-PLC- β 1	2.48 ± 0.29	3.78 ± 1.22
anti-PLC- γ 1	3.17 ± 0.45	13.08 ± 4.87
anti-PLC- δ 1	3.18 ± 0.47	12.41 ± 4.78
(b) control serum	2.14 ± 0.05	12.46 ± 0.40
anti-PLC- β 1 serum	2.05 ± 0.15	4.13 ± 0.13

a. Bovine liver plasma membranes (150 μ l of 0.5 mg ml⁻¹ in 25 mM HEPES, pH 7.5, 5 mM MgCl₂, 1 mM EGTA) were incubated on ice for 3.5 h with 1 μ g of the indicated monoclonal antibody or mouse IgG. Membranes (9 μ g protein) were assayed for 8 min in the presence of 2.8 ng GPA-42 or the appropriate buffer. b. Plasma membranes (150 μ l) at 1.25 mg ml⁻¹ were incubated with a 1/50 dilution of anti-PLC- β 1 serum or control serum (keyhole limpet haemocyanin-immunized rabbit serum) on ice for 4.5 h. Membranes (22 μ g) were assayed as in a. Results, which are expressed as μ mol phosphatidyl inositol 4,5-bisphosphate hydrolysed per min per mg of membrane protein, are means $\pm$ s.e.m. of two sets of three experiments. Bovine liver plasma membranes were prepared as described¹¹. Monoclonal and polyclonal anti-PLC antibodies were generated as described²⁵.

The α_i and α_s resolved from GPA-42 as well as $\beta\gamma$ subunits do not stimulate partially purified PLC¹¹ or pure PLC- β 1 (data not shown). Therefore, because α_0 and α_z are not expressed in liver¹¹, we investigated which G protein α subunits are represented by GPA-42. Novel α subunits from rat brain¹⁴ have amino-acid identity with members of a recently identified class of α subunits, specifically G α_q and G α_{11} (ref. 12). Because GPA-42 and these α subunits have similar M_s and cannot be ADP-ribosylated by pertussis toxin¹¹, we tested whether or not they were related. Purified GPA-42 was immunoblotted with various G protein-specific and common antisera, including antisera (WO82 and WO83) raised against peptides corresponding to unique sequences within α_q ¹⁴ (Fig. 3). Antiserum 588 (' α_{common} ' antiserum) detected no proteins in the GPA-42 preparation but antiserum 8645, directed against the most homologous region of α subunits, recognized a doublet of proteins in the 42–43K range (Fig. 3a). This doublet was also seen following silver staining (Fig. 3b). In addition, substoichiometric amounts of G protein β subunits and two contaminating proteins (36–38K), which were not normally present in GPA-42 preparations, were also detected. Antisera WO82 and WO83 both recognized the lower 42K band of the doublet, which we therefore believe to be α_q . Because the upper band of the 8645-detected doublet

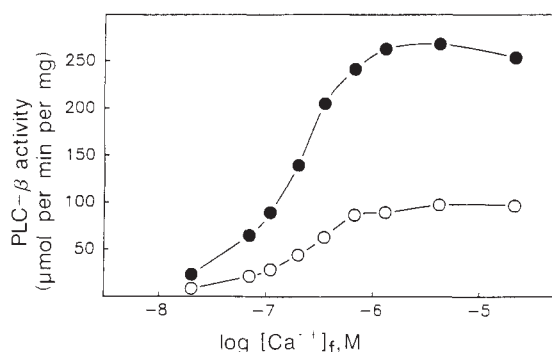


FIG. 2 The effect of Ca²⁺ on GPA-42 stimulation of PLC- β 1 activity. PLC activity was assayed with 2 ng PLC- β 1 in the presence of 5.6 ng GPA-42 (●) or the appropriate buffer control (○), over a range of Ca²⁺ concentrations. Free Ca²⁺ was maintained with Ca²⁺/EGTA buffer, with 2 mM EGTA, as determined with the COMICS program²⁸. Time of assay was 8 min. Results are from two experiments which were representative of similar experiments performed with another GPA-42 preparation.

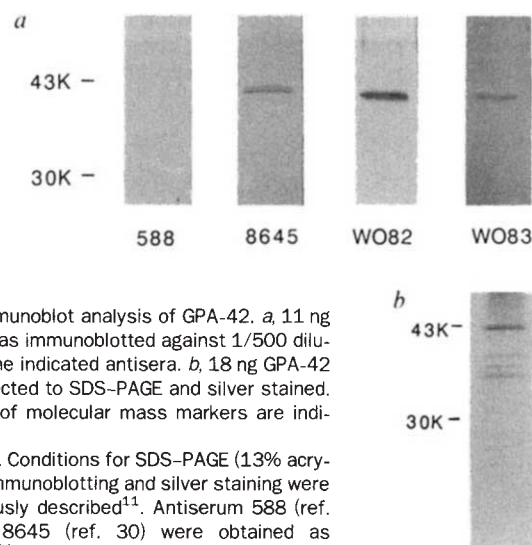


FIG. 3 Immunoblot analysis of GPA-42. a, 11 ng GPA-42 was immunoblotted against 1/500 dilutions of the indicated antisera. b, 18 ng GPA-42 was subjected to SDS-PAGE and silver stained. Positions of molecular mass markers are indicated.

METHODS. Conditions for SDS-PAGE (13% acrylamide), immunoblotting and silver staining were as previously described¹¹. Antiserum 588 (ref. 29) and 8645 (ref. 30) were obtained as described¹¹.

can cross-react with WO83 (data not shown), does not react with antisera specific for other known α subunits, and co-chromatographs with α_q , we believe it to be a closely related α subunit, possibly α_{11} (the WO83 peptide contains three amino-acid changes in α_{11}). We have recently resolved the two α subunits chromatographically and have shown that they both have PLC-stimulatory activity (manuscript in preparation).

Pharmacological and molecular approaches have increased understanding of many of the G protein-coupled receptors linked to PLC activation, but the G protein(s) and PLC species involved have until recently remained elusive. That the β 1 isozyme is a principal target of the pertussis toxin-insensitive stimulatory G protein of liver suggests that individual PLC isozymes mediate distinct signal transduction pathways. Three distinct isozymes of mammalian PLC- β have been identified³ but expression and reconstitution will be necessary to determine whether all three isozymes are targets of G-protein action. We cannot rule out the possibility that isozymes other than PLC- β 1 are also subject to G-protein regulation. A PLC purified from turkey erythrocytes that is subject to receptor and G-protein regulation in reconstituted erythrocyte ghosts^{15,16} may represent an avian homologue of mammalian PLC- β 1, although it does not cross-react with antibodies raised against the latter¹⁵. The data presented here suggest that PLC- α , a major isozyme in mammalian liver¹⁷, is not a primary site of G-protein action in liver plasma membranes. Because PLC- δ 1 is not activated by GPA-42 its role in signal transduction is unclear. In some cell types G protein-mediated PLC activation is pertussis toxin-insensitive, indicating the involvement of a distinct G protein. PLC- δ 1 may be an effector of such a G protein. A mutant cell line of Chinese hamster lung (CCL39) fibroblasts, which is defective in thrombin-induced phosphatidyl inositol turnover and mitogenesis¹⁸, possesses no immunoreactive PLC- δ 1. It is perhaps significant that thrombin-elicited phosphatidyl inositol 4,5-bisphosphate hydrolysis in CCL39 fibroblasts is inhibited by pertussis toxin¹⁹.

We show that members of the G_q class of α subunits are involved in PLC- β 1 regulation. Others have also observed stimulation of partially purified PLC activity by purified brain G_q in the presence of fluoroaluminate²⁰. Both proteins have widespread tissue distribution^{1,3,12} and homologues of each are present in *Drosophila* eye^{21,22}, so the pathway may be common to several eukaryotic tissues and species. Expression of the relevant complementary DNAs in appropriate cells should resolve whether one or more members of the G_q class activate PLC- β 1. Such an approach, perhaps involving PLC- β 1 coexpression, should show how disruption of this signal transduction pathway, such as by constitutive G-protein activation or overexpression,

may lead to disruptions of cellular function. Based on the recent precedents of *gsp* and *gip*^{23,24}, it will be of particular interest to determine whether GTPase-defective mutants of the G_q class possess oncogenic potential. □

Received 25 January; accepted 7 February 1991.

1. Rhee, S. G., Suh, P.-G., Ryu, S.-H. & Lee, S. Y. *Science* **244**, 546–550 (1989).
2. Majerus, P. W. *et al. Cell* **63**, 459–465 (1990).
3. Kriz, R. *et al. Ciba Fdn Symp.* **150**, 112–123 (1990).
4. Ullrich, A. & Schlessinger, J. *Cell* **61**, 203–212 (1990).
5. Wahl, M. I., Nishibe, S., Suh, P.-G., Rhee, S. G. & Carpenter, G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1568–1572 (1989).
6. Margolis, B. *et al. Cell* **57**, 1101–1107 (1989).
7. Meisenhelder, J., Suh, P.-G., Rhee, S. G. & Hunter, T. *Cell* **57**, 1109–1122 (1989).
8. Anderson, D. *et al. Science* **250**, 979–981 (1990).
9. Nishibe, S. *et al. Science* **259**, 1253–1256 (1990).
10. Boyer, J. L., Hepler, J. R. & Harden, T. K. *Trends pharmacol. Sci.* **10**, 360–364 (1989).
11. Taylor, S. J., Smith, J. A. & Exton, J. H. *J. biol. Chem.* **265**, 17150–17156 (1990).
12. Strathmann, M. & Simon, M. I. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9113–9117 (1990).
13. Carter, H. R., Wallace, M. A. & Fain, J. N. *Biochim. biophys. Acta* **1054**, 119–128 (1990).
14. Pang, I.-H. & Sternweis, P. C. *J. biol. Chem.* **265**, 18707–18712 (1990).
15. Morris, A. J., Waldo, G. L., Downes, C. P. & Harden, T. K. *J. biol. Chem.* **265**, 13501–13507 (1990).
16. Morris, A. J., Waldo, G. L., Downes, C. P. & Harden, T. K. *J. biol. Chem.* **265**, 13508–13514 (1990).
17. Crooke, S. T. & Bennett, C. F. *Cell Calcium* **19**, 309–323 (1989).
18. Rath, H. M., Fee, J. A., Rhee, S. G. & Silbert, D. F. *J. biol. Chem.* **265**, 3080–3087 (1990).
19. Paris, S. & Pouyssegur, J. *EMBO J.* **5**, 55–60 (1986).
20. Smrcka, A. V., Hepler, J. R., Brown, K. O. & Sternweis, P. C. *Science* **25**, 804–807 (1991).
21. Bloomquist, B. T. *et al. Cell* **54**, 723–733 (1988).
22. Lee, Y.-J., Dobbs, M. B., Verardi, M. L. & Hyde, D. R. *Neuron* **5**, 889–898 (1990).
23. Landis, C. A. *et al. Nature* **340**, 692–696 (1989).
24. Lyons, J. *et al. Science* **249**, 655–656 (1990).
25. Ryu, S. H., Cho, K. S., Lee, K.-Y., Suh, P.-G. & Rhee, S. G. *J. biol. Chem.* **262**, 12511–12518 (1987).
26. Ryu, S. H., Suh, P.-G., Cho, K. S., Lee, K.-Y. & Rhee, S. G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6649–6653 (1987).
27. Suh, P.-G., Ryu, S. H., Choi, W. C., Lee, K.-Y. & Rhee, S. G. *J. biol. Chem.* **263**, 14497–14504 (1988).
28. Perrin, O. D. & Sayce, I. G. *Talanta* **14**, 833–842 (1967).
29. Mumby, S. M., Khan, R. A., Manning, D. R. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 265–269 (1986).
30. Carlson, K. E., Brass, L. F. & Manning, D. R. *J. biol. Chem.* **264**, 13298–13305 (1988).

ACKNOWLEDGEMENTS. We are grateful to I.-H. Pang and P. C. Sternweis for antisera W082 and W083, D. R. Manning for antiserum 8645, and P. C. Sternweis and M. I. Simon for sharing manuscripts prior to publication. We thank R. Bryant and A. Ross for excellent technical assistance and J. Childs for preparation of the manuscript.

The molten globule protein conformation probed by disulphide bonds

Jonathan J. Ewbank* & Thomas E. Creighton*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

THE molten globule is a compact protein conformation that has a secondary structure content like that of the native protein, but poorly defined tertiary structure. It is a stable state for a few proteins under particular conditions¹ and could be a ubiquitous kinetic intermediate in protein folding². The extent to which native interactions, above the level of the secondary structure, are preserved in this conformation is not so far known. Here we report that α -lactalbumin can adopt a molten globule conformation when one of its four disulphide bonds is reduced. In this state, the three other disulphide bonds rearrange spontaneously, at the same rate as when the protein is fully unfolded, to a number of different disulphide bond isomers that tend to maintain the molten globule conformation. That the molten globule state is compatible with a variety of disulphide bond pairings suggests that it is unlikely to be stabilized by many specific tertiary interactions.

Two extreme but plausible models for the molten globule state are that it is either a collapsed, unfolded form or an expanded form of the native state that maintains the overall polypeptide backbone fold. NMR studies of apomyoglobin³ and guinea-pig α -lactalbumin⁴ have suggested that at least some

secondary structure and higher-order structure that is 'more or less native'³ persist in these molten globule states.

The most extensively characterized molten globule state is that of the α -lactalbumins (α LA), proteins of 123 residues in which the fully folded globular conformation is stabilized by a single bound calcium ion⁵. Without this ligand, bovine, human and guinea-pig α LAs exist in similar molten globule states at low ionic strength, intermediate denaturant concentrations, low pH or elevated temperatures⁴. In this state, the circular dichroism (CD) spectrum in the far-ultraviolet (peptide) region is similar to that of the native protein, but the spectrum in the near-ultraviolet (aromatic) region is almost absent¹.

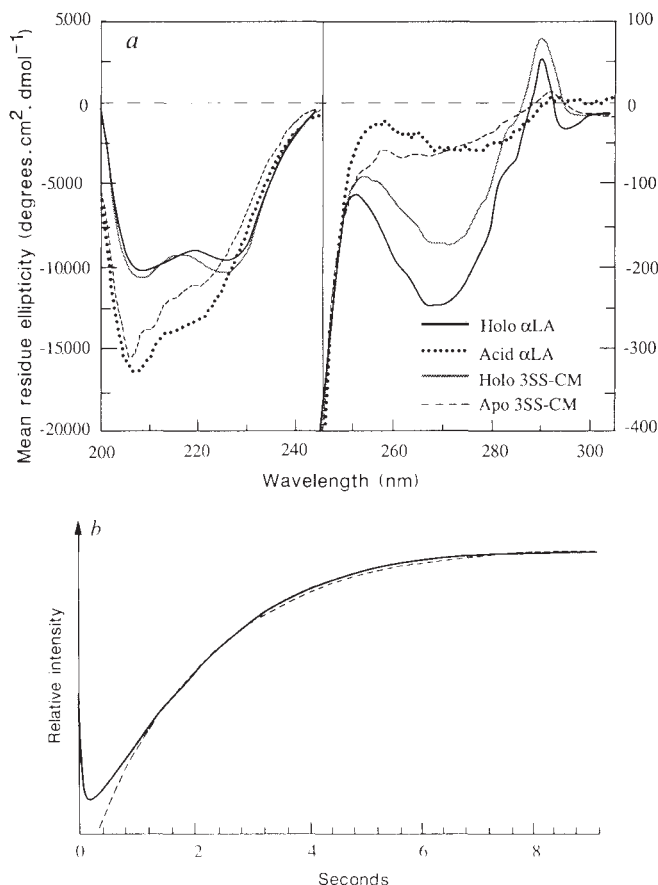


FIG. 1 Conversion of the partially reduced derivative of human α LA, 3SS-CM (with the 6–120 disulphide bond reduced and carboxymethylated), to the molten globule conformation. *a*, Comparison of the CD spectra of 3SS-CM in calcium-bound (holo) and calcium-free (apo) forms at neutral pH, with those of disulphide-intact α LA in the fully folded and molten globule states (at neutral and acid pH, respectively). *b*, Rate of conversion of 3SS-CM to the molten globule state upon removing bound calcium by addition of EDTA. The intrinsic fluorescence emission spectra of apo and holo 3SS-CM differ substantially, allowing the conformational transition to be followed. **METHODS.** *a*, α LA was obtained from Sigma and purified further by reversed-phase HPLC. 3SS-CM was prepared essentially as described by Kuwajima *et al.*⁹ and further purified by reversed phase HPLC. For the CD spectra, native α LA and 3SS-CM were dissolved in 0.1 M Tris-Cl buffer, 0.2 M KCl, 100 μ M CaCl₂ at pH 7. To obtain apo 3SS-CM, 0.8% (v/v) 0.5 M EDTA at pH 7 was added to 3SS-CM. For the acid form of α LA, holo α LA was titrated with 1 M HCl to pH 2.2. Spectra were acquired on a Jobin Yvon CD6 at 25 °C, with protein concentrations in the range of 8–15 μ M and cells of pathlength 10 mm and 1 mm. *b*, A solution of 2 μ M 3SS-CM in 0.1 M Tris-HCl, 0.2 M KCl, 100 μ M CaCl₂ of pH 7 was mixed in a stopped-flow apparatus with an equal volume of 0.1 M Tris-HCl, 0.2 M KCl, 5 mM EDTA, pH 7 at 25 °C. The intrinsic fluorescence was followed as a function of time, with an excitation wavelength of 280 nm and a 320 nm emission filter. The solid line is the average of ten transients and the dashed line is a single exponential with a half-time of 1.46 s.

* Present address: European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 6900 Heidelberg, Germany.

Disulphide bonds can be used as probes of protein conformation to characterize further the molten globule state. When disulphide bond partners may exchange, the most prevalent pattern of disulphide pairing will reflect the most stable conformation, as first demonstrated for insulin and ribonuclease A (ref. 6). Most native proteins possess a unique tertiary structure and a corresponding unique disulphide bonding pattern. The unfolded state of a protein is less defined; if its disulphide bonds are given the opportunity to exchange intramolecularly (for example, when unpaired cysteine residues are present), they will rearrange to give many different combinations, presumably those most favoured entropically.

The molten globule state of α LA is suited to this type of structural analysis. In its native state, one of the four disulphide bonds, linking Cys 6 and Cys 120, is hyperreactive and can be selectively reduced^{7,8}. In the presence of calcium, the conformation of the resulting molecule (holo 3SS) is stable and close to that of the native state⁹. Alkylation of its two free thiols does not alter the conformation⁹, but gives a chemically stable species (3SS-CM, when iodoacetic acid is used) which can be structurally characterized in the holo and apo forms.

Removal of the bound calcium, by the addition of EDTA, changed the conformation of 3SS-CM from that of the native state to that of a molten globule, as judged by characteristic changes¹⁰ in the CD spectrum. Apo 3SS-CM had a CD spectrum very similar to that of the well studied molten globule form of α LA with intact disulphide bonds that occurs at acid pH (ref. 1) (Fig. 1a). The intrinsic fluorescence of the protein was also altered, and this was used to show that conversion to the molten globule state occurred with a half-time of 1.5 s (Fig. 1b). Apo 3SS has a compact rather than an unfolded conformation, as indicated by its electrophoretic mobility when trapped with the neutral iodoacetamide, which is close to that of native α LA. Its electrophoretic mobility on urea-gradient gels¹¹ decreased gradually with increasing urea concentration (results not shown) consistent with the unfolding behaviour of a molten globule¹².

Unblocked 3SS has two free thiol groups as well as the disulphides 28–111, 61–77, 73–91. Consequently, there is potential for rearrangement of its intramolecular disulphide bonds and this can be used to address the question of how specific the interactions are that stabilize the molten globule state.

When calcium was removed from 3SS, its three disulphide bonds rearranged spontaneously, with a half-time of 3 min at pH 7 (Fig. 2a). The rate of rearrangement was the same as that observed for fully unfolded 3SS in 8 M urea (Fig. 2b). The rearrangement in the absence of urea could be reversed by the addition of excess calcium; that in the presence of urea additionally required dilution of the urea concentration to 2 M (data not shown).

The mixture of products obtained from disulphide bond rearrangement in the molten globule state is quite distinct from the mixture obtained under denaturing conditions, as judged by HPLC, CD and denaturing polyacrylamide gel electrophoresis (Fig. 3). In both cases most of the protein molecules retained three disulphides indicating that the rearrangement was largely intramolecular (Fig. 3d). The molten globule tended to aggregate, however, and even at the low protein concentration used here, a small proportion of products of intermolecular rearrangement were observed (Fig. 3c). The CD spectrum of the mixture of products obtained after the rearrangement of 3SS in 8 M urea is typical of an unfolded protein (Fig. 3b). The disulphides tended to rearrange in 8 M urea to those that increase the hydrodynamic volume of the protein (Fig. 3c). They are presumably between Cys residues close to each other in the primary sequence, as expected if the conformational flexibility is maximized under unfolding conditions. By contrast, rearrangement in the molten globule state tended to maintain the molten globule conformation, as shown by the far-ultraviolet CD spectrum of the trapped products (Fig. 3b), and the rearranged disulphide bonds were those that keep the polypeptide compact (Fig. 3c).

Bovine α LA behaved similarly, but had a greater tendency to aggregate and to undergo intermolecular disulphide rearrangement. These observations explain why reduced bovine α LA regains its native four disulphides and conformation only in the presence of calcium¹³.

The molten globule state of 3SS does not favour the native arrangement of disulphide bonds, but is consistent with a variety of disulphide bond pairings. Therefore, the molten globule conformation must be compatible with a number of different tertiary folds. The disulphide rearrangements occur as rapidly as when the protein is unfolded in 8 M urea. These observations suggest few, if any, of the tertiary interactions of native α LA are present in the molten globule state.

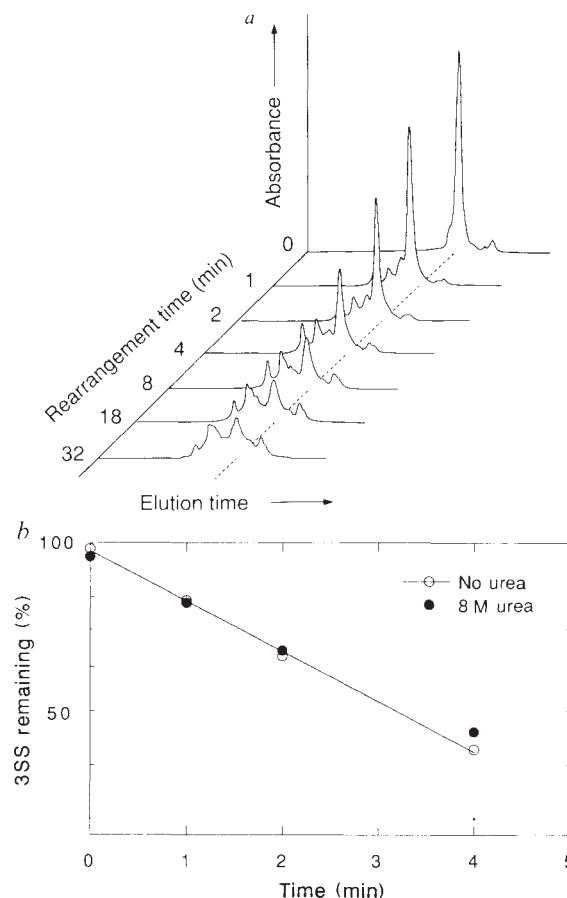


FIG. 2 Intramolecular rearrangement of the disulphides of 3SS. *a*, Rearrangement of 3SS in the molten globule state followed by reversed phase HPLC. *b*, Comparison of the rate of rearrangement of 3SS in the fully unfolded and the molten globule states.

METHODS. *a*, 3SS was generated by mixing 100 μ M α LA in 0.1 M Tris-HCl, 0.2 M KCl, 100 μ M CaCl₂, pH 7 with an equal volume of 400 μ M DTT in the same buffer. More than 95% of the protein was in the 3SS form after 10 min at 25 °C. The conformation was changed to that of a molten globule by a fivefold dilution of the partially reduced protein into 0.1 M Tris-HCl, 0.2 M KCl, 5 mM EDTA, pH 7 to remove the bound calcium. The reaction was quenched at the times indicated by the addition to aliquots of 1/4 volume of 1 M HCl. The aliquots were stored on ice until analysed by reversed-phase HPLC using a Dynamax-300A C4 column with a linear gradient of 25–50% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid. The dashed line indicates the elution position of 3SS. Identical profiles were obtained when Ca²⁺ was removed by EGTA rather than EDTA and when the final EDTA concentration varied over a 20-fold range (0.4–8 mM), indicating that the rearrangement is not caused by the chelating agent. *b*, For the rearrangement in the unfolded state, the conditions were as above, except that the partially reduced protein was diluted fivefold into 0.1 M Tris-HCl, 0.2 M KCl, 5 mM EDTA, 10 M urea, pH 7. Urea (8 M) does not alter the thiol-disulphide exchange rate¹⁶, but unfolds 3SS. The amount of 3SS remaining was calculated by integration of the HPLC chromatograms. The rate of rearrangement was greater at pH 8.7, as expected if thiol-disulphide interchange is rate-limiting.

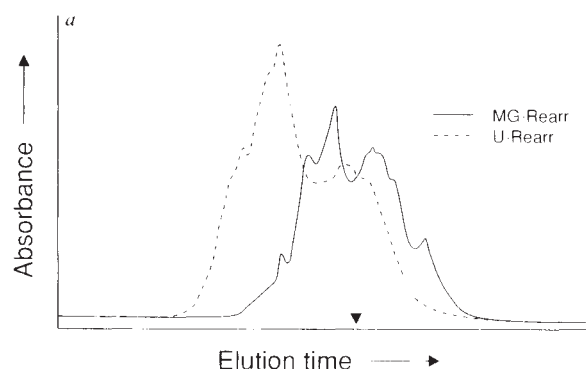
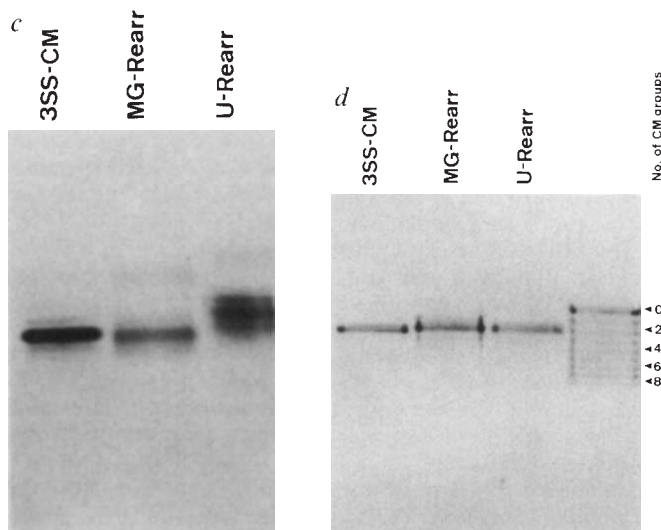
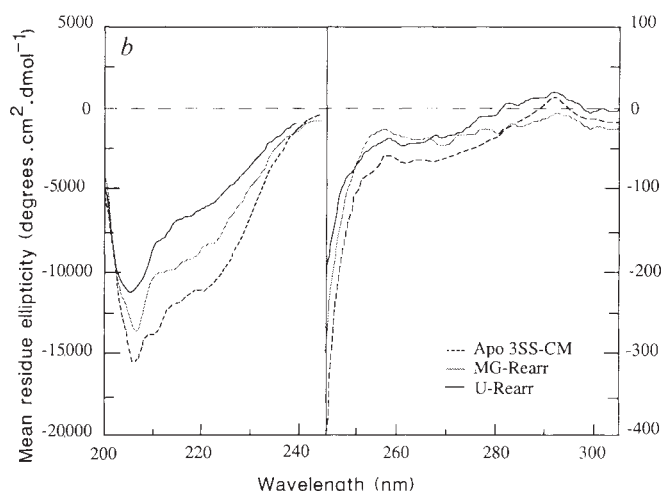


FIG. 3 Comparison of the iodoacetate-trapped products of rearrangement of 3SS in the molten globule (MG-Rearr) and fully unfolded states (U-Rearr). *a*, Analysis by HPLC. Rearrangement of 3SS in the two states was allowed to occur as in Fig. 2 for 15 min before trapping by the addition of 1/4 volume of 0.6 M potassium iodoacetate in 0.5 M Tris-HCl at pH 7. After 2 min at room temperature, the trapped material was stored on ice until HPLC analysis was performed as described in Fig. 2*a*. The arrowhead marks the elution position of 3SS-CM. The HPLC patterns of iodoacetate-trapped rearrangement products were similar to those obtained by acid trapping (Fig. 2). *b*, CD spectra. The trapped rearranged proteins obtained from the HPLC analysis described in *a* were lyophilized and dissolved in 0.1 M Tris-HCl, 0.2 M KCl, 4 mM EDTA, pH 7; the spectra were recorded as in Fig. 1*a*. *c*, Effect of disulphides on the hydrodynamic volume of the unfolded polypeptide chain. The hydrodynamic volumes of the rearranged proteins were measured by their electrophoretic mobilities in gels containing 8 M urea, 12.5% acrylamide and 0.3% bis-acrylamide; the high pH discontinuous buffer system of Davis¹⁷ was used. *d*, Number of disulphide bonds. The integral number of acidic carboxymethyl (CM) groups introduced by the trapping procedure was determined by the electrophoretic mobility (as in *c*) of the protein molecules after reducing all disulphide bonds present and blocking the thiol groups generated with the neutral reagent iodoacetamide. All the α LA molecules are converted to the fully reduced and alkylated form and differ only in their net charge, depending upon the number of CM groups present. The right-hand lane is fully reduced α LA that had been trapped with mixtures of iodoacetate and iodoacetamide, to generate a spectrum of molecules with from zero to eight CM groups¹⁸. Most of the rearranged molecules incorporated two CM groups upon trapping with iodoacetate, as did the original 3SS species, indicating that they retained three disulphide bonds. The minor other species probably resulted from further reduction of 3SS, and intermolecular



disulphide rearrangements that resulted from aggregation of the molten globule state.

It should be possible to define the α LA molten globule in greater detail by determining the disulphide pairings of the rearranged species. But some conclusions can be drawn simply from the substantial number of rearranged products. In the native α LA structure, one disulphide (28-111) is relatively exposed and close to Cys 6 and Cys 120 (about 15 Å away)¹⁴. Rearrangement of 3SS might be expected to occur most readily between the free thiols of Cys 6 or Cys 120 and the 28-111 disulphide bond, which could lead to only slight structural perturbation of the molecule's tertiary conformation. This limited rearrangement could generate a maximum of five new disulphide bond isomers. But many more than five isomers are generated by the rearrangements, although it is difficult to estimate the precise number as they are poorly resolved by the separation methods used (Fig. 3). Therefore, the two other disulphide bonds of α LA, which are close together on the opposite side of the native structure, near to the calcium binding site, must also be involved in the rearrangements. This conclusion is supported by the observation that most of the rearranged species no longer bind calcium: in contrast to native α LA (ref. 15) and 3SS-CM (data not shown), the mobilities of the trapped rearrangement products are unaffected by the presence of 10 mM calcium during electrophoresis under non-denaturing conditions. Consequently, the disulphide bond rearrangements of 3SS in the molten globule state must be accompanied by very considerable structural perturbation.

There is no single preferred conformation for the molten

globule state of α LA in the absence of fixed disulphide bonds. It favours many different compact tertiary folds, indicating that the molten globule state of α LA is much closer to an unfolded, but collapsed, form than to an expanded native conformation. □

Received 11 January; accepted 27 February 1991.

- Kuwajima, K. *Proteins* **6**, 87-103 (1989).
- Ptitsyn, O. B., Pain, R. H., Seminotnov, G. V., Zernovnik, E. & Razgulyaev, O. I. *FEBS Lett.* **262**, 20-24 (1990).
- Hughson, F. M., Wright, P. E. & Baldwin, R. L. *Science* **249**, 1544-1548 (1990).
- Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. *Biochemistry* **28**, 7-13 (1989).
- Kronman, M. J. *Crit. Rev. biochem. molec. Biol.* **24**, 565-667 (1989).
- Givol, D., De Lorenzo, F., Goldberger, R. F. & Anfinsen, C. B. *Proc. natn. Acad. Sci. U.S.A.* **53**, 676-684 (1965).
- Iyer, K. S. & Klee, W. A. *J. biol. Chem.* **248**, 707-710 (1973).
- Schechter, Y., Patchornik, A. & Burstein, Y. *Biochemistry* **12**, 3407-3413 (1973).
- Kuwajima, K., Ikeguchi, M., Sugawara, T., Hiraoka, Y. & Sugai, S. *Biochemistry* **29**, 8240-8249 (1990).
- Dolgikh, D. A. *et al.* *FEBS Lett.* **136**, 311-315 (1981).
- Goldenberg, D. P. & Creighton, T. E. *Analyt. Biochem.* **138**, 1-18 (1984).
- Pfeil, W., Buchkova, V. E. & Ptitsyn, O. B. *FEBS Lett.* **198**, 287-291 (1986).
- Rao, K. & Brew, K. *Biochem. biophys. Res. Commun.* **163**, 1390-1396 (1989).
- Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M. & Phillips, D. C. *J. molec. Biol.* **208**, 99-127 (1989).
- Thompson, M. P. *et al.* *Biochem. biophys. Res. Commun.* **157**, 944-948 (1988).
- Creighton, T. E. *J. molec. Biol.* **113**, 313-328 (1977).
- Davis, B. J. *Ann. N.Y. Acad. Sci.* **121**, 404-427 (1964).
- Hollocker, M. in *Protein Structure* (ed. Creighton, T. E.) 145-153 (IRL, Oxford, 1989).

ACKNOWLEDGEMENTS. We thank J. Foote for performing the stopped-flow fluorescence experiment; K. Kuwajima for unpublished results; K. R. Acharya and D. I. Stuart for the coordinates of baboon α LA; A. G. W. Leslie and P. J. McLaughlin for assistance with computer graphics; N. J. Darby for advice; and R. L. Baldwin, L. J. Berliner, C. Chothia, P. A. Evans, A. M. Lesk and O. B. Ptitsyn for discussions. J.J.E. received an MRC studentship.

The nomads of the sky

Edward Harrison

Comets: A Chronological History of Observation, Science, Myth and Folklore. By Donald K. Yeomans. Wiley: 1991. Pp.485. £24.95, \$38.95.

Like the red star that from his flaming hair
Shakes down diseases, pestilence, and war.
(Homer, *Iliad*.)

COMETS have fascinated the curious and terrified the superstitious since antiquity. In his authoritative *Comets: A Chronological History of Observation, Science, Myth and Folklore*, Donald Yeomans tells a gripping story in the history of science — of the comings and goings of comets, interwoven with biographical detail of the astronomers and mathematicians who have observed and studied them — that culminates in the rendezvous mission of the Halley spacecraft armada of 1986.

The opening chapters recount the popular and scholarly beliefs about comets through the ages. In observations and explanations, the Chinese were ahead of the Arabs and Europeans until the high Middle Ages. With the rise of the cartesian and newtonian systems, comets became a major astronomical problem: were they members of the Solar System or visitors from distant star worlds? Almost every astronomer performs a part in the colourful parade. Naturally, not everybody gets mentioned; a conspicuous omission by Yeomans is Otto von Guericke, Mayor of Magdeburg, celebrated for his sensational experiments with large evacuated vessels, who in 1650 argued that comets are periodic. In the appendix, Yeomans catalogues comets seen by naked eye in recorded history to AD 1700, in which Comet Halley can be traced back to the Chinese 'broom star' of 240 BC.

Until recent times, comets were feared as omens foreshadowing plague, famine and war. Disasters of one kind or another are forever with us and the credibility of the omens was rarely in doubt. Unintentionally, Edmund Halley added to the ancient fears a new fear of cometary impacts when he suggested in 1694 that an Earth-comet encounter had caused the biblical deluge. Prophets claiming that each bright comet foretells the end of the world have never been in short supply.

The immutable crystalline spheres of Aristotle and Ptolemy condemned all transient celestial phenomena as imperfect and sublunar. Comets in the Judaic-Christian-Islamic world of the Middle Ages were terrestrial exhalations, material enough to explain

their transience, ethereal enough to serve as vehicles of heavenly wrath. But science was on the rise, and observations began to indicate that comets exist above the lunar sphere. The realization in the sixteenth century that the crystal spheres would be shattered by the passage of comets contributed to the downfall of the ptolemaic system and the success of the copernican revolution. But fear of cometary apparitions cannot be blamed entirely on our aristotelian heritage; Homer regarded them as malefic agents, and cometary entries in the silk book of the Chinese, dating



The Earth is destroyed by a comet in this nineteenth-century French cartoon.

back to the fourth century BC, relate to significant events, usually disasters, such as rebellion and famine.

Yeomans condemns as astrological many of the celestial beliefs of the past. Most astronomers from Babylonian times to the late Middle Ages were astrologers. But astronomers generally fail to realize that astrology had not quite its present meaning. Astrology was the science of planets and stars, of celestial influences, dealing with the physical and etheric interactions between celestial bodies and the Earth. Eclipses and comets in the heavens and tides and seasons on Earth were associated with the play of astral forces. Side by side with astrology, particularly in the Roman Empire, flourished the temple arts of astrolatry (worship of astral bodies) and astromancy (astral divination and soothsaying). Since the seventeenth century, astrology has been superseded by astronomy, and all that is left to astrology is astromancy, concerned with the influence of planets on the

mundanities of human life. Astronomers justly condemn modern astrology, but go too far when they decry the astrology of the past; have they forgotten that out of the astral forces of the mediaeval universe arose the newtonian theory of universal gravity — a theory long rejected by continental scientists because of its astrological origins?

Yeomans makes an interesting point; he notices that Aristotle and Seneca relate their own views on comets after they have discussed and rejected those of their predecessors, "thus taking the remarkably modern stance of mentioning their sources only when they disagreed with them." This sweeping comment came to me as a shock; to what extent do modern scientists cover up their seminal sources, and give prominence only to references that are incidental and vulnerable to disproof? Is this one more feature on the dark side of science? We all travel on rutted roads of thought without paying tribute, but how many, when on little-trodden trails, conceal that they follow blaze marks made by others?

In the last chapters, Yeomans covers a wide diversity of cometary phenomena and emphasizes that we still do not know how comets originate. According to modern tentative ideas, the nucleus of a comet consists of a conglomeration of ice and dust and is only a few kilometres in size. When the comet approaches the Sun, the nucleus is surrounded by an extended atmosphere of gas and sub-micrometre grains of dust. Comet tails divide into two general groups: dust tails propelled by the pressure of sunlight, and gas tails entangled in the solar wind. A vast swarm (the Oort cloud) or approximately a trillion (10^{12}) comets

extends a fraction of the distance to the nearest stars; planetary and stellar perturbations and galactic tidal forces disturb this cloud and cause comets to plunge into the inner Solar System on highly elliptical orbits. Perhaps frequent cometary impacts four or more billion (10^9) years ago veneered the Earth with organic compounds essential for the origin of life; perhaps cometary impacts, on a much larger scale than the Tunguska event in 1908, have caused mass extinctions. No serious person now believes that living creatures inhabit comets, or that comet tails cause epidemics on brushing against the Earth (with the exception of F. Hoyle and C. Wickramasinghe who suggest that comets rain down on Earth the viruses that cause influenza epidemics).

Can the history of science be written in a scholarly and meaningful manner for readers who are not scientists? In Yeomans' book we see the past littered with discarded ideas. For an astronomer, the contrast in the ideas of

the past and the present greatly adds to the entertainment. But what of the reader unversed in astronomy? An editor once told me that she had no patience with the history of science; trying to understand modern ideas was difficult enough, why get confused and "misled by the false ideas of the past?" Authors of history-of-science books try to avoid confusing readers who are not scientists by retailing only ideas that modern science deems correct, and recounting only the lives of scientists that modern science deems important. These authors commit the whig sin of conforming the past to the present. Yeomans, to his credit, in a masterly work, tries to solve the problem; his first eight chapters deal with the history of comets, including the 'false ideas', and the last three chapters outline recent developments. But the going gets tough, and I still wonder, is it not true that the history of science, as a learned discipline, is for scientists only? □

Edward Harrison is in the Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01002, USA.

Cathedrals of science

Martin Ward

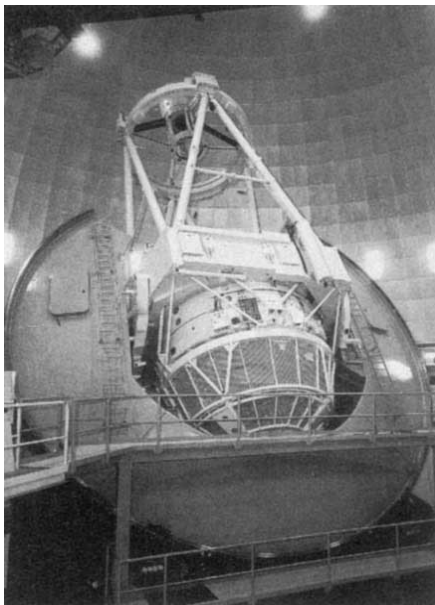
The Creation of the Anglo-Australian Observatory. By S.C.B. Gascoigne, K. M. Proust and M. O. Robins. Cambridge University Press: 1990. Pp.301. £40, \$59.50.

THE construction of a large modern telescope has been likened to the building of a mediaeval cathedral, in terms of the use of state-of-the-art contemporary technology and the aesthetic quality of the finished product. This point can be forcefully appreciated if one stands on the observing floor of the Anglo-Australian Telescope (AAT) and gazes up into the cavernous expanse inside the dome. *The Creation of the Anglo-Australian Observatory* by S. Gascoigne, K. Proust and M. Robins traces the history of the AAT from the early days of identifying the need for such a facility, through project management, technical design considerations, commissioning, and finally lists a selection of its major contributions to astronomy. Although the building of the AAT took considerably less time than the construction of a gothic cathedral, and was not interrupted by the black death, some equally traumatic problems had to be overcome.

The AAT is a major chronological landmark in the development of UK observational astronomy during this century. The book relates how it caused a fundamental change in attitude and observing practice from the previous style of pottering about with small telescopes located within Britain, at sites which are extremely poor by interna-

tional standards. A gentlemanly pursuit perhaps, but not one which resulted in Britain contributing greatly to the list of fundamental observational discoveries during the first part of this century. As our first major astronomical optical national facility, the AAT catapulted Britain into the big league. As a consequence, instead of joining the brain drain of the 1960s and 1970s, young British astronomers were able to look their Californian counterparts eye-to-eye without feeling like poor relations.

Much of the book deals with the mechanics of setting up the project, negotiating for funding from the respective British and Australian grant awarding bodies, and at a higher level, the political implications of a large-scale bilateral project. Although the book does go into some technical aspects of the design, both optical/mechanical and computational, its main purpose seems to be to record the organizational and personal struggles, and how by cajolery and compromise somehow the project was brought through to a successful conclusion. In consequence, the book will probably hold particular interest for persons involved in project management and organizational structure, rather than to those wishing to learn some astronomy. Indeed, it makes no claim to be a textbook.



The Anglo-Australian telescope, Siding Spring Observatory, Coonabarabran.

Some nice touches found in this book are the incidents of human interest, such as a senior member of the Australian National University (ANU) staff reputedly seen during the opening ceremony raising the university flag above that of the royal standard, in flagrant violation of protocol. Another is that we are given the opportunity to read the text of speeches by a prime minister and a royal personage, leading to interesting conclusions concerning their respective speech writers.

The authors draw some morals from their story. For example, as a first-class large telescope with state-of-the-art instrumentation

and a dedicated staff, the AAT produces scientific data of the very highest quality. These factors to a large extent overcome the disadvantage that the site chosen for the AAT, although probably the best available in Australia, is only in the second rank by global standards. Another conclusion is that the definition of the controlling body for the telescope should have been better defined at an early stage. This would have avoided a long running and often acrimonious dispute between the ANU and the AAT Board over who had the responsibility of running the telescope. This dispute blighted early attempts to appoint a director, and on occasion even looked like wrecking the project. Another important lesson concerns the basic design of the telescope. If a good design is available from another source, in this case the US project to build large telescopes in Chile and Arizona, there is little to be gained from re-inventing the wheel. Better to modify a sound design in the light of lessons learned or specific requirements, than to start from scratch. The latter approach would have been a recipe for extra expense and delay. The AAT was completed so quickly largely because it followed an existing design.

The book comments on some other features that contributed to the success of the AAT. For the first time at this level of sophistication, computers were closely interfaced with the operation of the telescope. This simple yet brilliant innovation resulted in the acquisition of astronomical objects, that is, the pointing of the telescope, with unprecedented accuracy. This gave the AAT its legendary reputation as a user-friendly telescope, a situation particularly valued by visiting astronomers. Finally, the symbiosis of the AAT and the nearby Schmidt Telescope used for wide-field survey work, is striking. Since 1988 the AAT and the Schmidt have been administratively combined to form the two-telescope Anglo-Australian Observatory. The book relates the curious tale of how the Schmidt Telescope was only funded originally because it fitted into a Science Research Council spend profile, thereby pushing radio astronomers aside, despite their having what was perceived at the time as a more desirable project.

The story of the AAT is certainly worth setting down, and a place should be found for *The Creation of the Anglo-Australian Observatory* in sections of libraries dealing with the history of science, project management, and general astronomy. Its publication is timely as it should serve to remind those who guide our scientific destiny that big science projects can pay back big dividends. In the case of the AAT this is fundamental knowledge about the universe. Returning to the analogy, as with cathedrals, big is beautiful too. □

Martin Ward is in the Department of Nuclear and Astrophysics, Oxford University, Particle and Nuclear Physics Laboratory, Keble Road, Oxford OX1 3RD, UK.

On target

R. V. Jones

Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance.

By Donald Mackenzie. MIT Press: 1991. Pp.464. £19.95, \$40.50.

A BOOK on inertial guidance by a sociologist might plausibly be assumed to deal with such matters as how to persuade a stubborn market to embrace some new product. It is therefore refreshing to find in *Inventing Accuracy* sociologist Donald Mackenzie surveying with authority and insight the development of the guidance systems that make intercontinental ballistic missiles (ICBMs) such a dominating threat.

A basic problem of ICBM guidance is how precisely can the 'brain' of a missile determine its position in flight, and hence guide it to the intended target. In free space the problem can in principle be solved purely from internal data provided by directional sensors and accelerometers in three orthogonal dimensions, with double integrations of acceleration data to establish displacement from the starting point. Such a system — which can be independent of radio aids and so invulnerable to radio countermeasures — was used in the German V-2 rocket, which used gyroscopes not only as direction indicators but also as accelerometers when mounted in pendulous suspensions which caused them to precess when accelerated. After 1945 these ideas were successfully developed in America for intercontinental missiles, thanks to superb efforts by the American gyroscope designer G. S. Draper; the Russians, too, took up the German work.

In practice, inertial guidance is complicated by the Earth's gravity and by Einstein's equivalence of gravity and acceleration in an isolated system; and physicists as renowned as G. Gamow and P. M. S. Blackett accordingly objected to the proposal that their effects could be disentangled. But because relevant data concerning the Earth's gravitational field could be pre-stored in a missile's 'brain' it proved possible to allow for the cumulative effects of gravity in computing the missile's position. Mackenzie describes how Draper won the argument with Gamow, and proceeded to make gyroscopes with the performance necessary to guide intercontinental missiles to within H-bomb range of their targets. Between 1947 and 1975 Draper reduced the drift rate of gyroscopes 100,000-fold to around 10^{-5} degrees per hour by inspired design and immaculate attention to detail. The ultimate design (post-Draper) was an electrostatically suspended free ball; this has, incidentally, been further developed for the Stanford Relativity Experiment where a drift rate of 10^{-7} degrees per year is the aim.

Besides the Draper-Gamow argument, Mackenzie also describes a controversy

between practical men and theorists on how many stars need to be observed for a position on the Earth's surface to be established as a check on inertial navigation. In principle two are needed, but it is possible to 'get away' with one if that one is known from the Nautical Almanack to be vertically over the target at the time when the missile should strike it. This 'Unistar' method depends, of course, on how accurately the true vertical can be established in the moving missile, and 'the problem of the vertical' is a further consequence of the Einstein equivalence.

Although thus dissecting some of the fundamental physical problems in inertial guidance, Mackenzie appreciates also the crucial need for ultra-cleanliness and skilled craftsmanship, quoting from the 1977 report of the Draper Laboratory:

We are concerned over a trend, a national one, that assumes that manual and mechanical skills are no longer necessary in a world populated by computers... that people who are skilled at making things that work are considered expendable... No computer ever made a piece of working hardware. No software ever made a measurement, ground a fitting, sealed a vacuum, cast a bearing, wound a core, heat-treated metal, magnetized a spoon, polished a lens, or etched a plate... if the pool of the highly skilled dries up, we would be seriously affected, working as we do in areas where craftsmanship is a principal ingredient.

Such was the skill involved that it "takes about six months and 300,000 dollars to rebuild an MX accelerometer" — to be used in triplicate in the MX missile to give a probable error of about 0.06 nautical miles at 6,000–7,000-NM range.

Although the MX, like Minute Man, was sponsored by the US Air Force, the US Army had long been in the ICBM field building on the German V-2 work with the Jupiter missile and the Saturn booster that launched the Moon shots, whereas the US Navy had Polaris and Trident. Mackenzie attends with insight to the rivalries and jostlings between the three US services, with the US Navy adopting Polaris despite its previous condem-

nation of the destruction of cities because it would give the Navy a say in national nuclear strategy.

There were rivalries and jostlings, too, between the proponents of the various types of gyroscope (including lasers) and accelerometers; and Mackenzie also points to differences in the US and Russian approaches, with the Russians rejecting the 'Unistar' technique, relying less on complex software, and preferring gas-bearings rather than fluid-flotation gyroscopes. It emerges, incidentally, that the MOM compass seemingly used by the Russians and made in Hungary is "the best commercial gyrocompass in the world", with the grinding of its ball-bearing raceways "superior to that available in the West at the time". Mackenzie remarks the special aptitude of women for the painstaking assembly of precision components on production lines, and also records that although in 1965 only 3.1 per cent of the scientists and engineers employed by NASA were women, the figure had risen to 10.2 per cent by 1987.

Inventing Accuracy is the product of thorough research into a wide range of aspects, and it provokes thought in many directions ranging from the technical (for example, a hypothetical 'Schuler' pendulum of length equal to the Earth's radius would have the same period as a satellite orbiting at ground level), to the human and political. In an epilogue its author concludes that "After a half century of 'thinking the unthinkable' — pondering nuclear holocaust — the time has surely now come to think the other unthinkable, a feasible world permanently free of nuclear weapons". And, we might add, all other weapons of devastation. Until then, we must cling to Winston Churchill's hope that "we shall by a process of sublime irony have reach a stage in this story where safety will be the sturdy child of terror and survival the twin brother of annihilation." □

R. V. Jones is at 8 Queen's Terrace, Aberdeen AB1 1XL, UK.



A rare photograph of a pygmy hippopotamus. Its limited distribution in West Africa makes it an endangered species. From *The Rainforests of West Africa* by Claude Martin (Birkhäuser, SFR 72, DM86).

General Chemistry

Donald A. McQuarrie and Peter A. Rock
Thoroughly revised and updated, this major textbook for mainstream chemistry courses, focuses on principles while interweaving the most important facets of descriptive chemistry. It is now available, for the first time, as an International Student Edition.
March 1991, 1,184 pages, 793 illus, board, £42.95; ISE, £19.95

Quantitative Chemical Analysis

Daniel C. Harris
Written in a clear, informal style that will appeal to students and instructors alike, this text is designed for students with a background in general chemistry who are preparing for work in related fields (such as the life sciences) or for advanced studies in chemistry.
March 1991, 758 pages, 419 illus, board, £42.95; paper, £21.95

Universe

William J. Kaufmann III
Thoroughly updated, the Third Edition of this text still offers an outstanding full-colour illustration programme, and has retained its highly acclaimed exposition, noted for its clarity and a sense of excitement.
January 1991, 656 pages, 710 illus, board, £27.95

One-dimensional and Two-dimensional NMR Spectra by Modern Pulse Techniques

Edited by Koji Nakanishi
This unique collection of selected spectra is a suitable guide for professional NMR spectroscopists and students. Using a picture book format and minimal theory, each example is accompanied by brief comments.
University Science Books
November 1990, 234 pages, paper, £29.95

The Hour of Our Delight

Cosmic Evolution, Order, and Complexity
Hubert Reeves
Hubert Reeves, one of France's most distinguished researchers and well-known popularizer of science, examines the evolution of our Universe, describing the increasingly organized structure of matter and our own species' climb up the ladder of complexity. *The Hour of Our Delight* won the prestigious Prix Blaise Pascal and is now published in English for the first time.
May 1991, 246 pages, 20 illus, board, £14.95

The Shadows of Creation

Dark Matter and the Structure of the Universe
Michael Riordan and David N. Schramm
Foreword by Stephen Hawking
More than ninety percent of the Universe is unseen, unknown; yet there is convincing evidence that it exists — as something called "dark matter". Join Michael Riordan and David Schramm as they explore this most remarkable mystery of the decade, describing efforts underway at observatories, laboratories, and particle colliders to determine the nature of dark matter.
April 1991, 278 pages, 62 illus, board, £15.95

Fractals, Chaos, Power Laws

Minutes from an Infinite Paradise
Manfred Schroeder
Fractals, Chaos, Power Laws reveals the fascinating consequences of new insights into the realm of symmetry. The reader will explore the surprising presence of self-similar symmetries in physics, chemistry, music, and the visual arts.
March 1991, 429 pages, 196 illus, board, £24.95

Discovering Enzymes

David Dressler and Huntington Potter
David Dressler and Huntington Potter convey the excitement of scientific detective work as they show how the attempt to understand alcoholic fermentation led to the discovery of enzymes, how the key to enzymatic function was uncovered, and how enzymes are isolated and studied at the molecular level.
Scientific American Library
January 1991, 263 pages, 150 illus, board, £16.95

The Science of Words

George A. Miller
The Science of Words offers both historical perspective and future prospects for psycholinguistics, while deepening our understanding of the relations among words, language, thought, and the human brain.
Scientific American Library
May 1991, 276 pages, 96 illus, board, £16.95

For details about the latest *Readings from Scientific American* please call 0865-726975 or write to the address below



W H Freeman

20 Beaumont Street, Oxford OX1 2NQ

Nonlinear explanations

Robert M. May

Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. By Manfred Schroeder. Freeman: 1991. Pp.429. £24.95, \$32.95.

ONE of my ambitions is to write an essay on why some books about science — Hawking, Penrose, Gleick — succeed beyond all reasonable expectation, whereas others languish. My problem is that, like the publishers themselves, I can explain these successes in retrospect better than in prospect. Easier to understand is the 'Jaws II' syndrome, whereby one book that has captured the popular imagination is followed by others that amplify the theme.

Which brings us to *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* by Manfred Schroeder. The general level and style suggest that Schroeder is targeting an audience of mathematically literate scientists, aiming to go beyond James Gleick and others both in dealing with a wider range of topics and in treating them in greater depth. Although the title leads off with Fractals, Chaos, much of the emphasis is on scaling phenomena or Power Laws and on their pervasive occurrence in physics, chemistry, biology, music and elsewhere.

I found the book very engaging, but also very exasperating. Schroeder covers a vast amount of territory, much of it going beyond what previous books have made familiar. There are, for example, interesting discussions of scaling and symmetry in relation to "acousticicians, alchemy, and concert halls", to Bach "composing on all scales", to trees, rivers, arteries, lungs, and stock markets. Old gambling paradoxes are discussed in relation to new themes. Some of the exposition is rather demanding, yet will reward those who persist: how to determine fractal dimensions from time series or fracture surfaces, or how to compute the dimension of a strange attractor.

The subtitle exemplifies the more exasperating aspects of the book. *Minutes from an Infinite Paradise* is a marvellous phrase, but I still have no clue how it relates to the book's contents. Far too many of the explanations are themselves nonlinear (you cannot understand any of it unless you understand all of it); some passages are excessively telegraphic, whereas others meander. For instance, there are 32 pages on the quadratic map, which go into some of the fascinating details of its cascades of period doubling and its ap-

■ *An Eye for Fractals: A Graphic and Photographic Essay* by Michael McGuire has just been published by Addison-Wesley at \$29.95, £24.75. It contains 90 black and white photographs plus an annotated bibliography.

parently chaotic behaviour. Despite the generous length, I found this discussion hard to follow, and the description of "chaos... interleaved with periodic windows" left me with the impression that the author does not really understand what is going on in this regime. Earlier popularizations of chaos have been criticized for giving insufficient credit to non-American workers. Schroeder discusses early Russian work, but does not mention the yet earlier work of the Finn Myrberg, nor of the French; given the abundance of references and other scholarly apparatus in this book, such omissions are less excusable than in earlier books.

At the telegraphic extreme is a chapter entitled 'Percolation: from Forest Fires to Epidemics', where the word epidemic is used in the third and the second to last paragraphs, but epidemics as such are never discussed. In the chapter on 'Noises: White, Pink, Brown, and Black', Schroeder refers to a recent paper by Redfearn and Pimm (see *Nature* 334, 613-615; 1988) as showing that fluctuations in terrestrial animal populations get larger as the observation interval lengthens. In fact, this paper has stimulated a lively correspondence, and its conclusions are by no means apodictic. Schroeder's point is, however, illustrated unambiguously by the pink noise spectrum characteristic of environmental fluctuations in the sea, which are not mentioned. These criticisms, and others that could be made, might look nit-picking, but — given the softness in the areas I know well — I am left wondering whether the book's admirably wide coverage has been achieved at the expense of consistent accuracy.

Schroeder was one of the first people to see the purely artistic possibilities inherent in computer graphics. Indeed, he won first prize at the International Computer Art Exhibition in 1969. The graphics in this book are thus excellent, and the inset of nine coloured figures would be stunning if earlier coffee-table chaos books and calendars had not already spoiled us. I especially like the illustration, and supporting analysis, of the fractal basins of attraction generated by Newton's iterative method applied to $z^3 = 1$. And those not already familiar with the Mandelbrot Set will enjoy the illustration of fat little Mandelbrots nested one within the other, an infinite regression that gives a kind of concreteness to pre-scientific visions of life not coded by DNA but rather unfolding from generation to generation through successive homunculi, nested like Russian dolls inside the egg.

More than other recent books in its general area, this book conveys a sense of the broad sweep of recent work on chaos, fractals and scaling laws. Read in this spirit, as an impressionistic guidebook and not as a map, it is fun. □

Robert M. May is in the Department of Zoology, Oxford University, South Parks Road, Oxford, OX1 3PS, UK.

More than meets the eye

Ian Stewart

Theories of Everything: The Quest for Ultimate Explanation. By John D. Barrow. Oxford University Press: 1991. Pp.223. £14.95, \$22.95.

Most scientists work in tiny corners of their subject, content to confine their activities to a few particular facets of the natural world. Others tackle the big problems that they believe lie behind these more mundane activities: the nature of the universe, its ultimate components, its origins. They search for a self-proclaimed 'theory of everything'; and they appear to be increasingly confident that they are about to find it.

Are they right? If they are, how much use will a theory of everything actually be? In his opening chapter, John Barrow remarks that ultimate questions had been falling out of fashion among scientists until recently. Suddenly thinking more appropriate to theology is showing up in the new theories of science, to the extent that theologians are having to come to terms with superstrings. "The theologians think they know the questions but cannot understand the answers. The physicists think they know the answers but don't know the questions. An optimist might thus regard a dialogue as a recipe for enlightenment, whilst the pessimist might predict the likely outcome to be a state in which we find ourselves knowing neither the questions nor the answers."

Any study of why science has returned to speculation on the grand scale, of which its speculations are, and of how they may affect our lives, must also take in many other matters of a philosophical nature. What is scientific explanation? In what sense does it capture the true behaviour of the natural world? Why is mathematics such a powerful weapon in scientific understanding? *Theories of Everything* is a beautifully written and thoughtful attempt to make sense of this difficult area of enquiry. It is worth buying for its apt and amusing quotations alone, such as Poul Anderson's "I have yet to see any problem, however complicated, which when you looked at it in the right way, did not become still more complicated."

What is a law of nature? The very word is a curious choice, for one cannot break a natural law. Laws are created by mankind to control human behaviour in an imperfect world. There is something rather anthropomorphic in the idea that nature, too, has laws and obeys them. But, without doubt, there is a strong belief, not just that the universe has patterns, but that it has a single unified pattern. We are dissatisfied with distinct theories for distinct realms of nature, however successful each may be: there is something fundamentally disturbing in the philosophical and mathematical inconsistency of quantum mechanics and general relativity. Inconsistency is an unhappy reminder that our 'natural' laws are man-

made: what better response than to unite everything in a single law?

The realm of quantum mechanics is that of the electron, that of general relativity is the galactic supercluster. At the tiniest scales, we find particle physics; at the largest, cosmology. But in modern cosmology the worm eats its own tail: the behaviour of fundamental particles is crucial for the first instants of the big bang. In this style of physics, it is impossible not to seek to reconcile the two theories, for, as logicians have always known, speculation within an inconsistent framework can produce any conclusion whatsoever.

Even within a framework of agreed laws, different initial conditions lead to different dynamical evolution. The initial conditions for our universe are probably unknowable, and this is somewhat awkward. Many cosmologists hope that the current state of the universe is largely independent of its initial conditions. Others seek to prescribe the initial conditions as part of the laws; or, bearing in mind that time itself may 'start' along with the embryonic universe, argue that the idea of an initial condition is simply misleading. Relativity complicates the question: causally disconnected regions of spacetime may have independent 'initial' conditions.

The role of initial conditions is very much bound up with the nature of time itself, something that a true 'theory of everything' must address. The same goes for the nature of matter. For instance, why are there so many identical elementary particles in the universe? All electrons are equal. Nothing on a macroscopic scale works this way; no two rocks, planets, clouds or bushes are exactly alike. But down on the tiniest scales, the universe is not as complicated as it might be. The symmetry principles of particle physics must surely be related to this puzzle; but are they a solution, or a consequence?

There is another crucial question that a 'theory of everything' must answer. Why do the various constants of nature (such as the speed of light or the mass of the electron) assume the particular values that we measure? For that matter, why are they constant? Indeed, are they constant? Here the anthropomorphic principle raises its head. Perhaps many of the 'constants' are arbitrary. But for human intelligence in its present form to be possible, they must take values very close to the observed ones. When the sole survivor of a plane crash asks 'why me?' the fallacy is obvious. But if there can be only one universe, the anthropic principle still tends to beg the question. The problem vanishes if we accept that (thanks to wormholes and inflation) there may be other universes. In one of

these a quasi-intelligent pseudoplasmod may be wondering why the speed of light, at 255 metres per hour, is exactly what is needed to make quasi-intelligent pseudoplasmods possible. Incidentally, the strong anthropic principle, which effectively holds us to be the purpose of the enterprise, looks less compelling when observed through the communicatory field-modulators of a quasi-intelligent pseudoplasmod.

Of such questions (though, I hasten to add, not that particular example) is *Theories of Everything* built. Its scope is, appropriately, vast; it draws from science, culture, myth and religion; and it touches upon many important concerns, ranging from chaos and symmetry-breaking to information-compression and the dilute wormhole approximation. Its intention is to talk about theories, rather than to explain them in detail, and this occasionally causes difficulties for the casual reader. Some assertions are questionable: for example that an "important lesson we have learnt from the mathematicians' approach to axiomatic systems is that one can quantify the amount of information that is contained in a collection of axioms. None of the possible deductions . . . can possess more information than was contained in the axioms." The spirit of the statement is correct, although I shall argue below that it tends to be misinterpreted, but its qualification through information theory is a red herring.

The author's main conclusion is not about the likely truth of the current attempts at 'theories of everything'. It is that, even if successful, such theories will not, in practice, render everything else obsolete. Yes, another anti-reductionist book. I tend to agree, though perhaps for different reasons. In practice, deductions from axioms — and laws — do 'contain' more information than the axioms themselves. More accurately, they make that information accessible. 'Knowing' laws is different from 'knowing' their consequences, even though the latter may well be implicit in the former. The quantification that is traditional in information theory misses the point here: there is a human cost in making the consequences explicit, and until this is done, the 'information' in the original laws is useless to us. This is why society pays mathematicians to make deductions that add zero information: it wants to know what the answers are, not just that they are implicit in the problem.

Similarly, physicists believe that the existence of crystalline states of matter is a consequence of quantum mechanics; but nobody can yet prove this mathematically. Until the question is decided, we do not know whether the 'information' in crystallography adds anything to that in quantum mechanics. But, whatever the answer, crystallographers will continue to use crystallography to study crystals. As Barrow so rightly says: "There is more to Everything than meets the eye." □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Naming names

John M. Charap

Moscow and Beyond. By Andrei Sakharov. Knopf: 1991. Pp.168. £14.99, \$19.95.

IN December 1986 Andrei Sakharov received a telephone call from Mikhael Gorbachev, recalling him to Moscow from his seven-year exile in Gorky to "resume his patriotic work". His *Memoirs* (published last year by Hutchinson), end with that call. By the early 1970s his was the clearest voice from within the Soviet Union, speaking with conviction, reason and passion for civil rights, for an end to the nuclear confrontation with the West, for an end to the rape of the environment in the Soviet Union. His opposition to the war in Afghanistan brought him exile by Brezhnev. Just three years after his recall to Moscow, Sakharov was dead. *Moscow and Beyond* is his account of those three years.

Back in Moscow in their small apartment, Sakharov and his wife Lusia (Yelena Bonner) received a stream of ambassadors, wellwishers and journalists. Lusia cooked for them, then took her turn to mop the landing of their communal building before typing his letters and press statements. Perhaps Gorbachev expected that these two ailing, elderly campaigners would leave it at that. But Sakharov took every opportunity to campaign against the continued detention of prisoners of conscience, stressing the inconsistency with the dynamic of *perestroika*. And he would name names, for it was always necessary to reinforce the general argument with particular individual cases.

Within weeks, Evgeny Velikhov invited him to take part in a Forum for a Nuclear-Free World to be held in Moscow in February 1987. Sakharov used that forum to elaborate his views on convergence — the rapprochement of the socialist and capitalist systems — and the nuclear disarmament which must go with it. Specifically, he criticized with reason and vigour the prevailing policy of the Soviet Union which linked negotiations on arms limitation to abandonment by the United States of the Strategic Defense Initiative (Star Wars); within two weeks the Soviet Union renounced this package principle with respect to intermediate range missiles (and later with respect to shorter-range and intercontinental ballistic missiles too).

During the Forum and in the following months, Sakharov met many western scientists, some old friends, others for the first time. But there is only scant reference to the substance of his scientific interests, and what there is betrays a nostalgia for when he was himself active in research. We learn little directly about his personal life, except for the deep love and companionship which Lusia brought to it. He was closer to her children

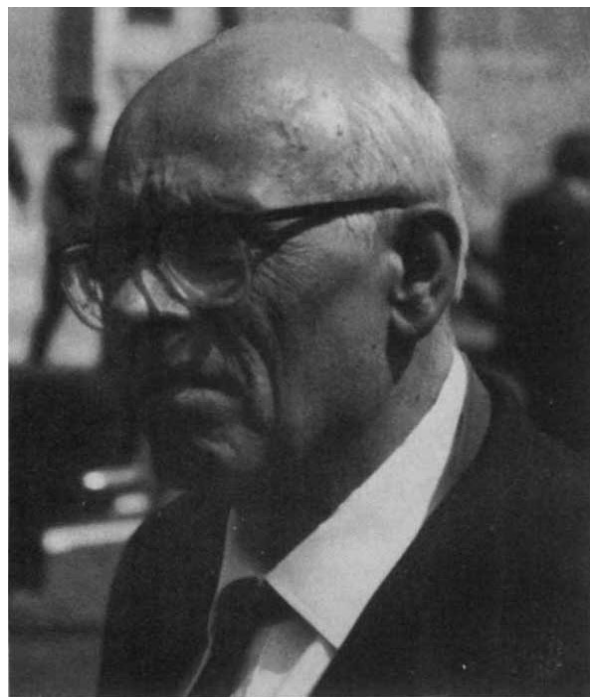
than to his own from his first marriage, and his elliptic, diffident mention of them is painful to read (as it probably was to write). Sakharov was more at ease writing about his public life which now became a whirlwind of activity.

Those last three years were crucial and testing years for *perestroika*. Sakharov was now a major participant, directly involved in the events which shook the country. Already in 1987 (unlike many in the West) he was not bedazzled in his appraisal of Gorbachev whom he saw as "trying to gain control of the political situation and strengthen his personal power by compromising with the forces opposed to *perestroika* instead of relying on democratic reforms. That's extremely dangerous."

Visiting Estonia in 1987, he recognized the example set by the Baltic Republics "with their popular movements for a genuine, not a fictitious *perestroika* and for a radical resolution of nationality problems through economic autonomy and the adoption of a new compact between the Union and its constituent Republics." Within weeks of the explosive eruption of interethnic violence in Nagorno-Karabakh in 1988 Sakharov had written an open letter to Gorbachev, and the fumbling efforts of the Kremlin to reconcile the Armenians and the Azerbaijanis was one of the first issues addressed by the Moscow Tribune, "a club, in effect an embryonic legal opposition" he had helped to found. Then on 7 December Armenia was shaken by an earthquake: Gorbachev flew back from the United States to visit the disaster area, only to be harangued by grieving people. On his way back to Moscow from his first trip abroad (he had met with Reagan, Bush, Shultz — and Thatcher — in New York, and was now a guest of President Mitterand in Paris), Sakharov immediately appealed for foreign assistance for the earthquake victims. He also discussed Nagorno-Karabakh with Mitterand, and *à propos* the fate of the Iraqi Kurds received an assurance that France would cut off military assistance to Iraq. Within days he was asked by Alexander Yakovlev, a Senior Secretary of the Central Committee and radical reformer among the party elite, to go to Armenia and Azerbaijan. He was appalled by what he found there, the corruption and ineptitude which frustrated the distribution of relief supplies, the plight of the refugees from both sides of the conflict, the stories of violence and torture. What the Moscow authorities were doing was too little and too late. Sadly, his advice seems to have made little impression on the authorities.

The book ends with the 1989 Congress of

People's Deputies, that extraordinary fortnight of turbulent debate under the eyes of the television cameras, the nation and the world. Sakharov had been elected (not without a struggle) as a representative from the Academy of Sciences, and from the start played a prominent and influential role. For him the central issue was constitutional reform. In the closing minutes of the Congress he proposed a Decree on Power (starting with repeal of Article 6 of the Soviet constitution, which ordained the Communist Party's monopoly of power — a proposal finally adopted in March 1990); proposed dramatic cuts in the armed forces; proposed "the creation of a new constitution based on horizontal federalism" which would give effective independence within the Union to the Republics; addressed the problem of the forcibly resettled nationalities (Crimean Tatars, for example); and stressed that "the most urgent political question is the confir-



Sakharov in 1989 continuing his patriotic work.

mation of the soviets and their independence . . . *All Power to the Soviets!*"

This was not empty rhetoric. In the brief time remaining before his death that December, he had drafted a new Constitution. He knew that he was leading an opposition, and in his last public address asked: "What is the meaning of political opposition? We simply cannot share responsibility for the actions of a government that is leading the nation to disaster and postponing the realization of *perestroika* for years to come . . . The only way, the only chance, for peaceful reform of the system is a radical quickening of *perestroika*." I would certainly go along with that. □

John M. Charap is in the Department of Physics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

New hopes, old doubts

Loren Graham

Man, Science, Humanism: A New Synthesis. By Ivan Frolov. *Prometheus*: 1991. Pp.342. £15.95, \$24.95.

IVAN Frolov, a leading philosopher in the Soviet Union and currently the editor of *Pravda*, has had a career that symbolizes both the new hopes and the old doubts about that country held by much of the rest of the world. Just as is the case with Gorbachev, Frolov rose to prominence on an anti-establishment reform platform. And just as with Gorbachev, he has now passed into a phase in which his work appears to be preserving the remnants of the old order more than continuing to push for a new one. But in the interim, Frolov, like Gorbachev, has helped to change the Soviet Union irretrievably. The story is a dramatic and complex one.

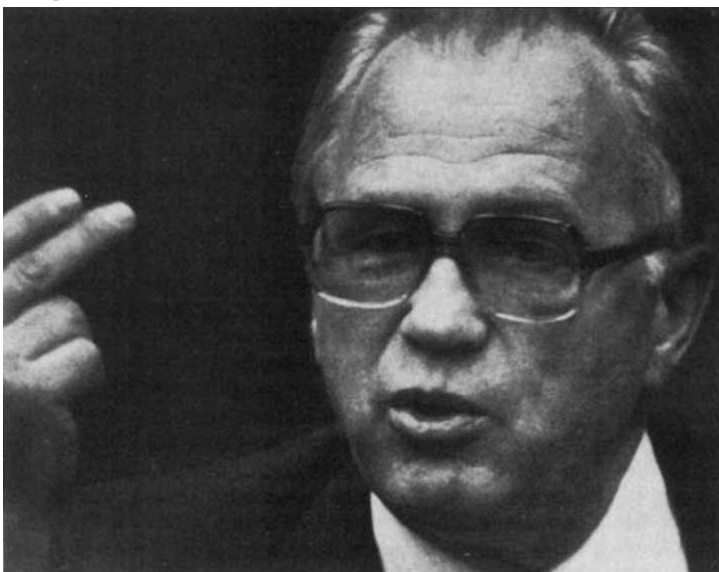
Frolov's main interests have always been the philosophy and ethics of science. In the 1960s he first came to the attention of intellectuals in the Soviet Union by writing a book in which he opposed the views of the pseudoscientist Trofim Lysenko, who still exercised great influence in Soviet biology. Frolov was a marxist, then and now, but he maintained that nothing in marxism required the denial of genetics. Indeed, Frolov observed that it was immensely ironic, and tragic, that the chromosome theory of inheritance, postulating the existence of material carriers of heredity, should be opposed by many Soviet marxists in the name of philosophical materialism.

Identified by this early work as a new and bold spirit among Soviet philosophers, Frolov went on to become the editor of the main philosophy journal in the Soviet Union, and there he set about refreshing Soviet philosophy of science by establishing closer ties with the natural scientists. At the same time, he became more and more interested in 'global' problems such as pollution of the environment, the shortage of energy resources, and biomedical ethics. The study of such problems was viewed by the old-fashioned ideologists as unorthodox; the dogmatists criticized the assumption made by most 'globalists' like Frolov that the problems of industrialized nations transcend class and economic rivalries to such a degree that the capitalist and socialist nations should overcome their traditional rivalries and work together.

With the decline of *détente* in the late

1970s, the militant ideologists opposed to Frolov's accommodationist views gained in influence. Frolov lost his job as editor of the main philosophy journal and was replaced by a much more orthodox thinker. Unable to continue his reforming efforts, Frolov turned more and more to studies of biomedical ethics, of which *Man, Science, Humanism: A New Synthesis* is one product. He is particularly interested in the morality of eugenics and genetic engineering applied to humans.

Frolov observes that Marxism is based on a relativistic vision of the history of civilization in which moral standards evolve in step with the development of material culture. Therefore, the possibility of the conscious and widespread application of genetic engineering to human beings in the future (even in ways that seem morally offensive now) can-



Ivan Frolov — since his editorship, *Pravda* has changed significantly.

not be excluded. On the other hand, maintains Frolov, it would be a great mistake to make such an effort now. For the time being, all eugenic ideas and all proposals to engineer a better human species should be rejected. The reasons, he says, are twofold: the science of genetics is still too incomplete and, even more important, power in the world is too unequally distributed. At some far future date, however, the question of improving the human species should be raised again; we should leave it to the people of that time, says Frolov, to decide the question, relying on what is anticipated to be their much better science of genetics within what will be, hopefully, a more just society.

Frolov's position contains a great deal of common sense. He avoids both the dangers posed by enthusiastic neo-eugenicists and those presented by fundamentalists who oppose all tampering with human genetics. But the weakness of Frolov's stand is its limited use in handling pressing practical questions. Frolov postpones the hard bioethical questions to the far future at a moment when all over the world physicians and researchers are already making decisions on questions

such as *in vitro* fertilization, the cessation of care to newly born deformed infants and terminally ill patients, the freezing of human embryos, the mapping of the human genome, the rules governing fetal research, the advisability of gene therapy, and the priorities among patients in receiving organ transplants. Frolov's highly abstract formulations, although on one level commendable, are not very helpful to working scientists and physicians facing these questions.

Frolov accepted Gorbachev's request that he become editor of *Pravda*, the main Communist Party newspaper, shortly before Gorbachev began to move in a more conservation direction. As a result, Frolov, who once was the main reforming spirit in Soviet philosophy, now must justify in his newspaper all the retrogressive steps recently taken by the Soviet government: the violent repression of

the Lithuanians last January, the tightening of control over Soviet television, the new restrictions on a free economy, and the effort to hold the Soviet Union together at all costs. Irritated by these editorial policies, and attracted by publications of the new independent groups in the Soviet Union, the Soviet reading public has massively turned away from *Pravda*. Its circulation is plummeting.

Even in this book Frolov seems more traditional than he earlier did. He warns that global problems should not be viewed outside the 'contradiction' between declining capitalism and victorious socialism; he praises the now defunct Council of Mutual Economic Assistance (the economic association of the East European states); he maintains that only in socialist and communist societies can new technologies be placed in the service of humankind.

Those of us who remember when Frolov fought for diversity and unorthodoxy in Soviet philosophy are sorry to see this conservative turn. Of course, even now all is not lost. Frolov is a true follower of Gorbachev, and like his leader, he has opponents on his right as well as on his left; both of them are far more enlightened than their possible right-wing replacements. Furthermore, in a journal which Frolov founded entitled *Chelovek* ('The Human Being'), many of his earlier ideas continue to appear.

Frolov's strength has not been moving from philosophical principles to practical action, as this book illustrates. Hence, no surprise that he is better at editing philosophy journals, or writing philosophical tracts, than at editing daily newspapers. □

Loren Graham is at the Massachusetts Institute of Technology, Program in Science, Technology, and Society, Cambridge, Massachusetts 02139, USA.

A is for atom

John Durant

Science Matters: Achieving Scientific Literacy. By Robert M. Hazen and James Trefil. *Doubleday: 1991. Pp.294. \$19.95.*

IN the midst of the Gulf War, US President George Bush took time out to greet the American Association for the Advancement of Science (AAAS), gathered in Washington for its annual meeting. Having pointed out that the US budget included substantial funding increases for mathematics and science education, President Bush added that "All sectors of society must recognize the importance of scientific literacy and strive to achieve it".

Scientific literacy is a buzz phrase in American educational circles. It stands for what the general public ought to know about science, and its widespread use reflects concern about the performance of the American educational system. In 1987, the English Professor E. D. Hirsch Jr captured this concern in his best-selling book: *Cultural Literacy: What Every American needs to Know* (Random House, 1987). Hirsch argued that the unity of American culture depended upon a common stock of generally shared knowledge, which he listed in the form of some 5,000 essential names, phrases, dates and concepts covering the entire world of learning.

In drawing up his list, Hirsch called upon the services of the physicist James Trefil. Trefil took advice from colleagues about the key concepts and facts that were "truly essential to a broad grasp of a major science". The result was several hundred items, ranging from 'absolute zero', through 'mutation', 'nuclear fission' and (amazingly) 'ontogeny recapitulates phylogeny', to 'Y chromosome'. In 1988, the list was expanded into a *Dictionary of Cultural Literacy* (Houghton-Mifflin, 1988). Starting out from the same concern "to provide only the constellations of basic facts and concepts that you need to understand the scientific issues of the day", Trefil and Earth scientist Robert Hazen attempt to review the world of science in 18 synoptic chapters. The first chapter is entitled 'Knowing', and deals with the idea that the universe is regular and predictable. Then come chapters on everything from energy and electricity to evolution and ecosystems.

Reviewing all of the natural sciences between one set of covers is a daunting task, but Hazen and Trefil approach it with relish. Throughout, they concentrate on key ideas and principles and avoid technical details. They use clearly headed sections to organize their material, and from time to time they highlight fundamental principles in large bold print. For example, chapter 5 on quantum mechanics is introduced with the

emboldened statement that, "Everything — particles, energy, the rate of electron spin — comes in discrete units, and you can't measure anything without changing it".

I share Hazen's and Trefil's enthusiasm for equipping the general public to make sense of the scientific issues of the day, I admire their synoptic grasp of the natural sciences, and I recognize their considerable skill in conveying the broad outlines of science in clear, nontechnical prose. Nevertheless, I have two major reservations about *Science Matters*. First, the book reflects its authors' disciplinary backgrounds in being too heavily weighted in favour of the physical sciences (14 chapters) and against the life sciences (four chapters). Even worse, it contains virtually nothing about the single area of science that is unquestionably of most direct relevance to the public: medical science.

Second, *Science Matters* is almost entirely taken up with a summary of the well-established findings of science. It devotes just one page to the scientific method, and just three pages to scientists. The account of scientific method is woefully inadequate, whereas the discussion of scientists is largely taken up with bizarre caricatures of the people who work in different disciplines. Thus, "For some reason many physicists, particularly those in universities, seem to enjoy appear-

ing sloppy and disheveled — always the ones without ties at faculty meetings"; whereas biologists are "the only group of scientists who routinely wear white lab coats".

Caricatures such as these do nothing to assist the general public in coming to terms with science; on the contrary, they merely perpetuate the myth that scientists are very peculiar people. But more serious even than this is the fact that Hazen and Trefil present science as a vast edifice of unproblematic knowledge built up by the routine application of 'the scientific method'. This view ignores the methodological diversity of science, as well as the difficulties that arise from scientific ignorance, uncertainty and controversy.

Science Matters is supposed to enable nonscientists to make sense of the scientific issues of the day. But the scientific issues of the day are real-life questions in which real-life people debate matters that are technically and/or socially controversial. Because it ignores this real world of scientific practice, *Science Matters* provides only a part of what nonscientists need to achieve scientific literacy. □

John Durant is at the Science Museum Library, Exhibition Road, South Kensington, London SW7 5NH, UK.

Articles of faith

George Marsden

Can Scientists Believe? Some Examples of the Attitude of Scientists to Religion. By Sir Nevill Mott. *James and James: 1991 Pp. 182. £21, \$40.*

FOR about a century in western civilization, roughly from the 1860s to the 1960s, the belief became widely held that scientific thinking was at odds with Christianity. Although, as a number of the authors of *Can Scientists Believe?* point out, this notion rested on a dogmatic conception of scientific knowledge that is now generally considered passé, the idea of an inherent conflict between science and religion has persisted.

In fact, of course, many scientists believe in Christianity and other religions. In the West this is particularly true in the highly religious United States, where over a fourth of scientists are active in churches and half describe themselves as religious. These percentages are not as high for the American public generally; but contrary to some popular belief, scientists are considerably more likely to be religious than are social scientists or humanist academics. So, even though rates of religious practice are much lower in most other western nations, the title question of this collection *Can Scientists Believe?* has an empirical answer.

Nonetheless, interest persists in this volume's main theme of how believing scientists relate their faith. Sir Nevill Mott's own resolu-

tion of the issue arose from being reassured by a vicar that one could be active in the Anglican Church without literally believing its creeds. But only a few of the other 14 contributors to this volume share the Cambridge physicist's strategy of resolving the problem by broadening the definitions of Christianity. Rather, most reflect on the implications of their scientific knowledge for a rather traditional faith, Protestant, Catholic or in one case Jewish. V. Ya. Frenkel, Professor of the History of Science, Leningrad, is the one contributor who takes a very dim, virtually comtean, view of religion. He remarks that he only ever met one highly intelligent man, a mathematician, who was religious. (The contrast to the number of religious scientists in the United States suggests the degree to which such issues are not decided merely on an intellectual basis.)

Wide diversity of opinion is the most striking general feature of the essays by believing scientists. No reader could be happy with all the contributions because a number contradict each other (as on whether the anthropic principle is of merit). Also a few are little more than impressionistic personal testimonies. Nonetheless, many others are valuable reflections on the relation of science to faith. So almost any reader interested in the topic is likely to find enough to make this a worthwhile volume.

One recurring theme is that science and theology represent two complementary ways of looking at reality. As Richard Bube of Stanford puts it, theology asks "who?" and science asks "why?" Several also note that scientific knowledge is nonetheless a lot

more like theological knowledge than is commonly recognized. For instance, C. W. Francis Everitt, also of Stanford, points out the similarities in being dependent on a community and dealing with insoluble mysteries. A number of authors argue that scientific truths are too often revised to provide dogmatic refutations to religious claims. For instance, Christopher Moss points out that recognition of the subjective elements of science, conditioned by social circumstances have tended to desecralize science. Knowledge of God, Moss adds positively, has the moral seriousness of justice at its core.

Arguments for free will versus a materialistic determinism provide the most recurrent theme. Several authors suggest (though one dissents) that indeterminacy in quantum mechanics has a bearing on the question of free will. Statistician, D. J. Bartholomew, in summarizing his study, 'The God of

Chance', provides some ingenious additional arguments. For instance, he suggests that randomness is not incompatible with design because there are still highly probable aggregates. Moreover, a 'random search' may be an efficient way for a mind to operate, although allowing room for the free will. Sir John Eccles argues that demonstrable influences of minds on brains substantiates free will.

Books on the beliefs of scientists constitute a venerable genre to which this one is a worthy contribution. The essential message of this volume is that science is not legitimately an obstacle to religious belief. Perhaps now we need a sequel, *Can Humanists Believe?* □

George M. Marsden is in the Department of History, University of California at Berkeley, Berkeley, California 94720, USA.

Heroines of science

Janet Browne

The Scientific Lady: A Social History of Women's Scientific Interests, 1520-1918. By Patricia Phillips. Weidenfeld and Nicolson: 1990. Pp.279. £25, \$35.

THE presence — or more usually, the absence — of women in science has been the focus of much concern during the last decade or so. Academics and teachers go out of their way to encourage female pupils to consider a career in the subject, perhaps explaining that equal opportunities, though not perfect, have never been better: Victorian mores, they say, have passed at last. *The Scientific Lady* will therefore come as something of a surprise. The central theme of Patricia Phillips's book is that women's scientific activities were at their zenith during the eighteenth and nineteenth centuries and have steadily deteriorated since. Science, it is claimed, has not opened its doors willingly to women since the Taunton Commission of 1864-8.

According to Phillips, the first recognizably modern science as formulated in the seventeenth century was a low-status activity: natural philosophers were thought to be little more than artisans, their experiments requiring mechanical skills, their theories depending on physical demonstrations of effects. By contrast, the study of classics and the ancients were the highest point of intellectual life. Where women would only rarely aspire to studying Latin and Greek at that time (although some discussed by Phillips plainly did), the low status of science made it more permeable to female enthusiasts. Women like Mary Astell, Marie de Gournay and Margaret Cavendish (Duchess of Newcastle) found it possible to create a niche for themselves and worked hard in the field.

With the great social changes of the eighteenth century there came a popularization of science among the European gentry which further stimulated female participation in scientific activities. This was the age in which Caroline Herschel studied astronomy, women wrote children's textbooks, and Aphra Behn translated Fontanelle's influential conversations into English.

"Pray Sir," said one of Fontanelle's heroines, "leave Adorers alone, and let us speak of the Sun." The way women took over the media of science is an important aspect of the wider cultural context of the eighteenth century and is well laid out here. In particular, one little-known facet of the diffusion of science among women is brought to light by Phillips: printed diaries of the period seem to have sometimes included scientific questions expressly intended for female readers (perhaps a precursor to the Christmas quiz). The answers were promised in the next year's issue. A series of manuscript letters to the publishers conveying the answers well in advance shows just how educated some of these ordinary, otherwise unknown women were.

By the nineteenth century, women had carved a capacious niche for themselves, although Phillips seems rather to overstate the case with headlines like "The female science student was a familiar sight." Nevertheless, she is right to emphasize the diverse and often ignored functions taken up during that period. It is good to have attention drawn to the great numbers of women who listened to Humphry Davy and Michael Faraday lecture on their researches at the Royal Institution: fashionable occasions though these were, the audience was primarily there to learn.

Equally welcome are the (too brief) discussions of the careers of significant figures like Augusta Ada Lovelace, Mary Somerville and Elizabeth Garrett Anderson. The thrust of these later chapters is the proposal that as science became more prestigious — and as emphasis on a classical education waned — men pushed women out of their former niche. Once science was transformed into an important profession women's participation dramatically decreased.

Phillips's argument is a powerful one that bears careful attention, although some of her sweeping judgements may seem unlikely to professional historians of science and a few aspects surely overemphasized for effect. It is hard to think of Sir Isaac Newton as a rude mechanical, for example, although this could apply to Robert Hooke. More could have been said about French *salons* in which intellectual women reigned supreme. How odd not to find Florence Nightingale somewhere in the text: her impact on nineteenth century women and on culture at large cannot just be ignored. The analysis is furthermore predominantly confined to European ladies of high social standing. Chapters dance about through the centuries and on two or three occasions appear to contradict themselves. Sloppiness is seemingly only just kept at bay,



The great reformer — Elizabeth Garrett Anderson and her husband in 1870.

ready to break out at any moment: the attractive dust-jacket, for instance, reads differently to the title page. But what this book misses in weight is balanced by a breezy, readable overview of the topic. Patricia Phillips is an anecdotal author with considerable style. It is social history with few pretensions and can be recommended to any general reader. □

Janet Browne is at the Darwin Archive, University Library, Cambridge CB3 9DR, UK.

Where do we go from here?

Stephen Lock

Strangers at the Bedside: A History of How Law and Bioethics Transformed Medical Decision Making. By David J. Rothman. *Basic Books*: 1991. Pp.320. \$24.95.

IN 1964 the World Medical Association produced an important statement on human experimentation, now known after the city where the meeting took place. The Helsinki Declaration was itself an updated version of the 'Nuremberg Code' formulated soon after the postwar revelations of the experiments in the Nazi concentration camps. Despite the code, however, by the late 1950s reports in medical journals were suggesting that ever more nontherapeutic research was occurring without the informed consent of the patients concerned. The new declaration emphasized that such consent was vital, and in Britain official bodies quickly produced guidance documents. But three years later the focus in Britain sharpened considerably: in *Human Guinea Pigs* (Routledge, 1967) a book expanded from a 1962 magazine article, Maurice Pappworth, a London physician, argued that the ethical problems arising from human experimentation had become crucial. Giving references documenting a large number of studies in children, pregnant women, prisoners, the dying and the old, among others, he pressed for two developments. First, research should be carried out only in true volunteers, and, second, one vital safeguard for patients would be the return of the 'physician-friend'.

Like all whistleblowers, Pappworth encountered much opposition from his fellow professionals, although the debate produced results. Every UK health authority established a research ethics committee and a new society of medical ethics set up groups in every medical school, and a journal and bulletin. And the appearance of advice documents and the wider debate have continued, with Ian Kennedy's 1980 Reith Lectures. *The Unmasking of Medicine* (Unwin Hyman, 1981), being followed by more professional hand wringing and sharpening of the issues. Elsewhere also there have been similar debates and solutions, particularly in the Nordic countries.

Regrettably little of this is recorded in David J Rothman's *Strangers at the Bedside* — although he concludes that possibly in the

United States the revelations at the Nuremberg trials were perceived as irrelevant. Thus his book is virtually — though not explicitly — limited to events in the United States. This lack of an international dimension diminishes an authoritative and beautifully written book. My reaction is not that of the Czechs in 1938 — bawling Neville Chamberlain's dismissive "far off country of which [Britons] know little." Rather, an examination of the debates and differing solutions elsewhere might have teased out the possible reasons for them and thrown more illumination on the questions Rothman discusses. For example, with their comprehensive system of good general practice, Britain and Denmark now have what Pappworth proposed as one solution: a physician-friend for the patient — whereas in Sweden patients have unrestricted access to a specialist. Yet of the three countries it is Denmark that has experienced the greatest public furor over ethical issues, and one can only speculate on the reasons.

But, this said, Rothman's account is first class. It reads well, as befits an account by a historian who is also a professor of social medicine. His subtitle, *A History of How Law and Bioethics Transformed Medical Decision Making*, suggests his theme: the mid 1960s to 1970s were critical in the evolution of bioethics. During this time doctors' discretion became increasingly circumscribed by institutional review boards, Food and Drug Administration regulations, and a Patients Bill of Rights while other professionals appeared — lawyers, philosophers, and sociologists. As a result both the style and the substance of decision making have been transformed, and the patient has a new friend in the bioethicist — a relationship, Rothman points out, that now crosses class barriers.

One man who started it all, Rothman's protagonist, Henry Beecher, was the third driver of the founding troika (the others being Pappworth and Hugh Clegg, editor of the *British Medical Journal* and behind the Helsinki Declaration). In 1966 an article in the *New England Journal of Medicine* by Beecher, a Harvard professor of anesthesiology, recorded 22 examples of ethically dubious human experimentation. Even before its appearance the account was an obvious hornet's nest. Beecher's previous presentation at a conference excited media attention and also led two Harvard colleagues to refute his views at a press conference. And initially the *Journal of the American Medical Association* had rejected the article, with two adverse referees' reports, while, on the advice of legal colleagues, Beecher omitted the references documenting these studies (printed in Rothman's book for the first time).

There were other reasons why concern over human experimentation surfaced at this time. For Rothman is on firm ground in showing that in the 1960s all professions came under threat, the thalidomide tragedy

destroyed illusions about continuous progress in medicine, and the reception of Rachel Carson's *The Silent Spring* (Hamilton, 1962) testified to a deep public suspicion of science. And a further major reason was a perceived change in doctors themselves. Until 1945 — many had practised little differently from their earlier fictional counterparts — George Eliot's Lydgate, with his professional pride, and Chekhov's Dr Askov exhausted from combating a typhus epidemic. After the war the generalist, a well-liked local figure, became reluctant to do house calls and, with the dominance of Medicare and Medicaid, was seen to have become preoccupied with his bank balance. Indeed, the generalist all but disappeared, to be replaced by an impersonal specialist working with frightening machinery in a distant institution and having scientific kudos as his priority.

The screw of public suspicion was turned a whole revolution by alarm over organ transplantation. Unlike patients with kidney disease, where dialysis was a feasible alternative, those with some forms of heart or liver disease had no alternative treatment. The need to obtain organs for transplantation from cadavers led to an obligation to define brain death, an agonizing process emphasized in Britain by the furor over a British television programme querying whether organ donors were really dead. And increasingly in the United States clinical dilemmas came to be decided not by doctors but by ethics committees or the courts. Thus in a series of lawsuits over noncompetent patients, the courts went against the relatives' wishes, ordering, for instance, the surgical treatment of a newborn with severe malformations and the continuance of artificial ventilation in a young brain-dead woman.

David Rothman's book largely ends in the mid 1970s, although in an epilogue he emphasizes the valuable role of the US president's commission for the study of ethical problems. Given that this could have issued guidance in the new ethical debates, it is regrettable that President Reagan was to discontinue it. For recently the thrust of these has shifted to other issues such as surrogacy and research in the pre-embryo, the human genome, and genetically manipulated organisms, not to mention priorities in allocating scarce resources, as the Oregon senate has recently been engaged in doing. Many countries have set up national bodies to consider such issues and in Britain we await the outcome of proposals for such an organization discussed in April 1990 at a Nuffield Foundation conference. Yet again international comparisons are instructive, an approach that Rothman should be persuaded to adopt in future editions as well as to continue his valuable researches into the present day. □

Stephen Lock is at the British Medical Journal, BMA House, Tavistock Square, London WC1H 9JR, UK.

■ Two books of archaeological interest are *The Passage of Arms: An Archaeological Analysis of Prehistoric Hoards and Votive Deposits* by Richard Bradley (published by Cambridge University Press, £30, \$44.50) and *The Vikings* by Else Roesdahl (Allen Lane, hbk £17.99, \$26.95; pbk £5.99, \$7.95).

Renaissance man

H. Christopher Longuet-Higgins

Models of My Life. By Herbert A. Simon. *Basic Books/Sloan Foundation Science Series*. 1991. Pp.448. \$25.95.

RIDDLES, puzzles, problems — these are the stuff of Herb Simon's life, and of his autobiography, the latest in a distinguished series sponsored by the Sloan Foundation. The quizzical title, matching those of his earlier books *Models of Man*, *Models of Discovery* and *Models of Thought*, challenges the presupposition that a single interpretation could do justice to a whole life; and the cryptic dedication "To Dorothea, so aptly named" hints at the deep happiness that has supported his wide-ranging accomplishments.

Simon's two best-known contributions to human thought are the theory of bounded rationality, for which he was awarded the Nobel Prize in economics, and his joint work with Allen Newell and Cliff Shaw on the foundations of artificial intelligence (AI). (The theory of bounded rationality highlights the boundary conditions within which economic agents maximize their expected utility — not least their incomplete knowledge and imperfect rationality.) Simon's principal metaphor for his life is that of someone exploring a maze, making a decision at every turn; but this is also his model of the administrator making imperfectly informed decisions, of the chess player struggling against the clock, or of the computer engaged in selective search guided by heuristics. Someone able to perceive such unities might well justify description as a Renaissance Man; but Simon prefers to present himself as a mere polymath, with a number of parallel roles and interests:

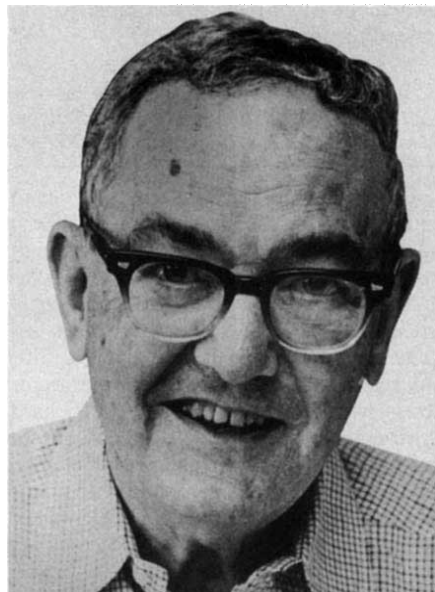
Perhaps clearest is the theme of the scientist and teacher . . . seeking the Holy Grail of truth about human decision making. In my case even that thread is woven of finer strands: the political scientist, the organization theorist, the economist, the management scientist, the computer scientist, the psychologist, the philosopher of science.

To these roles he adds that of the private person, family man, university politician and science adviser, and asks which of these will step forward at the call for 'the real' Herbert Simon. But by and large the book is mercifully free from Holy Grails and rhetorical questions, and Simon writes best when he is at his least self-conscious.

The book is divided into four panels (the metaphor is that of the triptych, or rather 'tetraptych'). The first portrays a mainly idyllic childhood, describes the author's undergraduate education at the University of Chicago, and closes on his twenty-first birthday in 1937. A formative experience was his attempt, with a group of teenage friends, to tame an area of wilderness and stock it with Hereford cattle. The cattle had

ideas of their own and broke down every fence the pioneers constructed, to get out of the pasture. "No doubt", writes Simon, "my later deep skepticism of the *a priori*ism of mainstream economics had some of its origins in this experience."

At university Simon switched from economics to political science. But as for international politics, "few of us spend any large part of our lives addressing the questions of peace. In my own case, I know one of the reasons: the feelings of frustration and futility that are aroused when I try to think about problems that have no visible solution." He



Herbert Simon — a puzzling life.

attended mathematical biology classes given by N. Rashevsky, whom he judged "rather cavalier in his attitude toward data", as a result of which he was "generally ignored by biologists". He claims that a respect for real-world data has always headed his list of scientific virtues:

. . . in empirical science the final test is not mathematical elegance or *a priori* plausibility, but the match between theory and data. I certainly learned that lesson somewhere, ultimately overcoming my innate Platonism and armoring myself against the aesthetic lures of neoclassical economics, so responsive to mathematical elegance and so indifferent to data.

The art of polemic is evidently not lost. But science would be a lot easier, and much duller, if 'the data' were never consistent with more than one theory!

The second panel describes 'The Scientist as a Young Man'. His first job was with the City Managers' Association, permitting a taste of research into 'the science' of admin-

istration. To most academics, administration is a mere chore, to be wistfully contrasted with research or teaching; but Simon succeeded in making a virtue of necessity. "I learned early that (in principle, at least) it gets easier, not harder, to administer as you move upward in an organization." As for teaching: "How does one instill curiosity into people? Without it, this 'educational' process is a continual struggle between a student who is trying to get by, and a teacher who is trying to catch him at it . . ." One of Simon's tasks at Illinois Tech was to teach courses on urban land economics and city planning administration to senior architects. There is a blistering attack on the architectural profession:

Architects are notoriously prone to design buildings that are bid in at 40 percent over the agreed budget. And they are notoriously inept with such prosaic details as air-conditioning, energy efficiency, waterproof roofs, and all manner of things their clients and the inhabitants of their buildings consider important . . . No society is affluent enough to provide its architects with all the steel and glass and concrete they need to save their artistic souls.

The rebellious streak that surfaces here and elsewhere may have been partly responsible for the loyalty investigations to which its owner was occasionally subjected.

The rest of the second panel describes Herb Simon's move to Carnegie Tech in 1949 to help found the Graduate School of Industrial Administration, and the intellectual developments that finally culminated in the birth of artificial intelligence. The year 1955 was a watershed in Simon's scientific career. Until then a fairly conventional economist/political scientist, he transferred his empirical studies from the real world to the psychology laboratory and began to couch his theoretical ideas in the formal languages that are used for programming computers. By the end of that year Newell, Shaw and Simon had invented list-processing computer languages and had used them to create the Logic Theorist, a non-numerical program that took less than five minutes to find an elegant proof of proposition number 2.85 in *Principia Mathematica*. Russell's response to the news —

I am delighted by your example of the superiority of your machine to Whitehead and me. I quite appreciate your reasons for thinking that the facts should be concealed from schoolboys. How can one expect them to learn to do sums when they know that machines can do them better?

— might have turned the head of the most hard-boiled researcher:

Put less technically, if more boastfully, we invented a computer program capable of thinking non-numerically, and thereby solved the venerable mind-body problem, explaining how a system composed of matter can have the properties of mind.

The remainder of the book freewheels gently down from the intellectual summit, taking in

the author's career in scientific politics, his Nobel prize and other honours — "I soon learned that one wins awards mainly for winning awards" — and various foreign adventures, including a memorable visit to Beijing that coincided with the Tiananmen Square uprising, thus flouting his newly enunciated travel theorem:

Anything that can be learned by a normal American adult on a trip to a foreign country (of less than one year's duration) can be learned more quickly, cheaply and easily by visiting the San Diego Public Library.

(My own travel theorem states that most of the world's troubles arise from people visiting places where they do not belong and interfering in matters that are not their business.)

The book is a mine of shrewd observations: on the need of students for their teachers' affection, and vice versa; on the roles of analysis and synthesis in science — the latter much neglected; on the inexactness of the 'exact' sciences, at least in their application to practical problems, and on a common stumbling-block to interdisciplinary research:

I would not give a dollar to assist a typical political scientist to collaborate with a typical economist unless each one of them gave me a sworn statement that he would study seriously . . . the discipline of the other for at least a year.

As a historical document, the autobiography covers a period of astonishing intellectual ferment: the explosion of the information sciences after the Second World War, the resulting transformation of psychology by the information processing paradigm and the use of computer simulation for getting a grip on the behaviour of highly complex systems. Particularly welcome is Simon's support for the view that one working AI program is worth a hundred opinionated articles on the powers and limitations of machines or networks of one kind or another.

Models of My Life achieves most admirably the goal of the Sloan Foundation to "make the scientific experience and the excitement of discovery accessible to the general reader." □

H. Christopher Longuet-Higgins is at the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Nature or nurture

Peter Bryant

The Origins of Exceptional Abilities. By Michael J. A. Howe. Blackwell: 1990. Pp.262. £35, \$39.95.

THE lives and achievements of talented people raise some interesting questions for psychologists, who would like to know the reasons for their extraordinary skills and would also like to find out if it is possible to teach more people to be gifted. These two questions are related, of course: if gifted people acquire their talents because of the experiences that they had during their childhood, it ought to be possible to provide other children with the same kind of experiences. The trouble is that it is usually very difficult to work out what are the causes of each gifted person's great success. Many gifted people work extremely hard at practising and improving their talents, and often, as children, receive a great deal of encouragement to do so from their parents and teachers. But we usually cannot tell whether these apparently propitious circumstances are the cause or the result of the person's talent. The young Mozart, for example, practised his music a great deal and he received much coaching from his father, but he also came from a musical family and it is quite possible that genetic factors played a considerable part in his extraordinary talents. It is also possible that he worked as hard as he did and received so much attention and encouragement because, for genetically determined reasons, he was unusually talented. We have no way of disentangling cause and effect.

This means that anyone who sets out to convince us that the reasons for exceptional talents are environmental ones has a difficult task on his or her hands. Yet this is the position that Michael Howe adopts in *The Origins of Exceptional Abilities*. He does acknowledge from time to time that it is difficult to prove that the particular circumstances in which Darwin, Mozart, Sir Richard Burton and several others were brought up were responsible for their remarkable gifts, but he seems to be sure that the environmental explanation is the correct one. He writes,



The child Mozart — was he born great or was greatness thrust upon him?

or course, at some length about the lives of these individuals and of several other people with exceptional talents, but he also marshals research on ordinary people to support his case.

His argument about the lives and particularly the childhood of exceptional people is unconvincing. His claim that they lived in unusual environments and had unusual experiences could only have any weight if it could be shown to be true. But we cannot rule out the possibility that other, less successful people were brought up in similar circumstances, and even if that were certain we would still have the chicken-and-egg problem of being uncertain whether the circumstances caused the talent or vice versa. The life histories of these extraordinary people always makes interesting reading, and they form the more interesting part of Howe's book, but the stories are not new and their relevance to our notions of intelligence have been discussed in more detail in books by Gruber, Gardner, Feldman and Treffert.

The work that Howe reviews on ordinary people is in the end no more persuasive. His main concern is to show that it is possible to influence people's skills by changing their experiences. Children, he argues, can be taught to read earlier, their motor skills can be improved, and their acquisition of language can be enhanced. All this is true but it has no obvious bearing on people with exceptional talents. Successful programmes for teaching reading do not produce exceptional readers — just children who can read. At times too, Howe seems to treat other people's views as hard evidence, provided that he shares those views. For example, he states that "Vygotsky showed that human cultures, internalised by children largely through the mediation of social interactions, not only provide access to skills and knowledge but also determine the fundamental sign systems that underlie knowledge". This is certainly what the Russian psychologist thought to be the case, but he never 'showed' that his theory was right, and in fact did very little in the way of empirical research.

Howe's charitable assessment of those who take the same position as his own contrasts strongly with his approach to other views with which he disagrees. "Far too often", he writes, "people have accepted uncritically beliefs about the importance of this or that factor simply because they seem plausible, or appear to be true or are widely considered by others to be true." But, disappointingly, he does not go on to identify these 'people' or to tell us exactly what their uncritical beliefs are.

The Origin of Exceptional Abilities raises an important and interesting question and provides a firm, but in the end quite unconvincing, answer. In fact no definite answer is possible at the moment. Until one is, we will need, and we still have to wait for, a fair and reasoned account of the evidence for and against all the plausible hypotheses about the reasons why some people acquire astonishing talents and the rest of us do not. □

Peter Bryant is in the Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK.

Both eminent and neglected

David Knight

William Whewell: A Composite Portrait. Edited by Menachem Fisch and Simon Schaffer. Oxford University Press: 1991. Pp.403. £47.50, \$45.

WILLIAM Whewell was one of the great pundits of nineteenth-century science. President of the British Association, twice president of the Geological Society, and then master of Trinity College, Cambridge, he was a meritocrat whose career shows the social mobility possible for the able scientist. He actually coined the word 'scientist', as well as 'anode', 'cathode' and 'ion', and the geological categories 'uniformitarian' and 'catastrophist'.



William Whewell (1794–1866) – 'science was his forte and omniscience his foible'.

With his treatises, he might almost be said to have invented history and philosophy of science as academic subjects. He also wrote on moral philosophy and theology: in a famous quip by Sydney Smith, science was said to be his forte and omniscience his foible.

Given his central position in the intellectual world of the mid nineteenth century, it is surprising how quickly he was forgotten: perhaps because he stood out like a Canute against the rising tide of specialization. It is very good that we should have in *William Whewell: A Composite Portrait* a series of essays about him, because they illuminate the theory and practice of the science of his day, and because some of his thoughts are still relevant to understanding scientific method.

His most famous books were the *History and Philosophy of the Inductive Sciences*

(1837, 1840); in which in opposition to John Herschel, John Stuart Mill and 'baconian' orthodoxy he proposed that good science consisted of authentic facts ordered by the active mind; which gave to each science its appropriate 'fundamental idea'. This determined its particular level in the hierarchy: chemistry required more than mechanical principles, and biology more than chemical ones. The unity of the sciences for Whewell would not be achieved by the reduction of them all to mechanics or some other basis; it was a matter of sharing a common inductive method.

The word 'science' in Whewell's time embraced any definite body of knowledge, and he hoped to see his sophisticated version of the inductive process making political economy and moral philosophy genuinely scientific. He believed that speculative and deductive thinkers, D. Ricardo in economics and utilitarians such as W. Paley and J. Bentham in ethics, had gone badly astray; as had P. S. Laplace in seeing no need of God in his deterministic cosmos. As a teacher, he distinguished 'permanent' subjects like geometry and classics as being the best discipline for young minds, who could come on to 'progressive' sciences like chemistry later; and he was thus a conservative in curriculum reform at Cambridge.

A composite portrait is an interesting idea, and on the whole it works extremely well: like a series of spotlights following an actor, the various authors illuminate him from different angles, and they have mostly got him in focus and written clearly. There is of course some overlap, and some disagreement: over whether Whewell can be described as a liberal Anglican, for example, and whether Darwin's *Origin of Species* is a good example of Whewell's 'Consilience of Inductions' or not.

There is still room for a biography, because there are gaps here; and there still seems to be a problem about Whewell's reputation and the reception of his ideas. Although he was so celebrated, his philosophy was not taken up until the twentieth century, when he was seen as a precursor of Popper, and when physics has very different fundamental ideas from those of 1850. It may be that he lived too long, surviving as a living fossil into the darwinian era; or conversely that his ideas were ahead of his time. But to be both eminent and neglected is indeed curious. □

David Knight is in the Department of Philosophy, University of Durham, 50 Old Elvet, Durham DH1 3HN, UK.

Coffee-table enzymology

Richard Perham

Discovering Enzymes. By David Dressler and Huntington Potter. Freeman: 1991. Pp.264. £16.95, \$32.95.

"WHAT we want is a story that starts with an earthquake and works its way up to a climax." David Dressler and Huntington Potter, authors of the beautifully produced *Discovering Enzymes*, approvingly quote Samuel Goldwyn and clearly set out to follow the injunction: no other book on enzymes that I know begins with the Big Bang and six pages and several billion years later, can be found rhapsodizing on the part played by enzymes in "[guiding] molecules across the threshold of life". Heady stuff, but accompanied by some exquisite colour photographs: a rock formation, a seashore, a developing embryo, a Jan Breughel, a collection of neurons (or is it nature imitating art – a Jackson Pollock? No, it is neurons), a blood clot, a silk moth emerging from a cocoon.

We then take a step back and are given an animated history of the discovery of enzymes. This too is richly illustrated (Jacques-Louis David's celebrated portrait of Antoine Lavoisier and his wife, for example – no enzymologist works like that today!) and is full of interesting personalization (von Baeyer's sour comment on Eduard Buchner, his former pupil, who crucially discovered cell-free fermentation, "This will bring him fame, even though he has no chemical talent". (Buchner was awarded the Nobel Prize for Chemistry in 1907.) The concepts of specificity, the lock-and-key hypothesis and the identification of enzymes as proteins (not forgetting ribozymes) are handled well. But I am not persuaded that it was wise to limit the discussion of enzyme structure and reaction mechanism almost entirely to chymotrypsin and related enzymes. The serine proteinases have an honourable place in past and present enzymology but it is strange to find no mention of lysozyme, of the haem proteins, of kinases or dehydrogenases, of folding domains or quaternary structure. Chymotrypsin is used to good effect in illustrating many features of enzymes, and the text conveys the excitement that unravelling the structure and function of a protein can bring, but we are denied a fuller picture of the extraordinary (and beautiful) range of structure-function relationships in enzymes. It also eliminates reference to coenzymes and cofactors, without which many enzymes cannot function and no metabolic pathway can exist.

Where the extended discussion of serine proteinases does take us is into the biochemistry of blood clotting and fibrinolysis. This allows an introduction to enzyme cascades and signal amplification and the potential for enzyme therapy. To conclude,

the authors take up the enzymology of signal transduction in the nervous system. The failure to deal earlier with ATP-linked reactions is mitigated here in the description of the sodium-potassium ion pump. Alzheimer's disease, nerve gases, the McNaghten Rules, are all brought into a rousing climax.

Discovering Enzymes can safely be recommended to anyone, undergraduate upwards, who wishes to become acquainted with the current excitement in the field of molecular enzymology, though it lacks the breadth of coverage to fit that bill on its own. One or two small quibbles: the strength of an

ionic bond is not the same as the force between two point charges; the velocity-substrate concentration plots, as drawn, are not rectangular hyperbolae; and it will sow confusion if amino-acid residues in proteins are referred to as subunits. Perhaps it is best summed up as that miracle of rare device, a coffee-table book on enzymes; it should fulfil a useful purpose in presenting a major area of the life sciences in a colourful and attractive way to a wider public. □

Richard Perham is in the Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK.

Modern shamans

Alison Jolly

Walking with the Great Apes. By Sy Montgomery. Houghton: 1991. Pp.280. \$19.95.

JANE Goodall, Dian Fossey and Biruté Galdikas have transformed our knowledge of chimpanzees, orangutans and gorillas — and with it our view of women's place in nature. In her joint biography, *Walking with the Great Apes*, Sy Montgomery traces three extraordinary lives. But she does more than tell their story; she tries to show that these three women are not scientists but modern shamans who reach out with an emotional current to the wild. She reveals the empathy, even the love, one must feel to spend decades with wild creatures, but is she right that science and love exclude each other?

The book has lively, lucid writing, and is constructed as a three-part fugue. Montgomery balances tact and openness, and respect without adulation. To me she too often knows what her subjects were thinking 20 years ago, or just what emotion is mirrored in an ape's eyes, but mostly this is a fine read.

She brings out well the differences between the three.

Jane Goodall's "westernness stands out like a porcelain tea-cup on a rough-hewn tree-stump". To judge from her story, though, Goodall's true mettle must be more like high-strength titanium, especially now in her crusade for the humane treatment of chimpanzees that have fallen into human hands. Galdikas makes her home in the swampy Bornean rainforest she compares to "the original Garden of Eden". She lives with her Dayak Indonesian husband, two of her children, a dozen or more ex-captive orangutans, Dayak guides, American Earthwatch volunteers and Indonesian graduate stu-

dents. As for Dian Fossey, Louis Leakey said years before her murder, "Her life was a tragedy, and will always be a tragedy". She was tragic in the Greek sense, not as a person to pity, but as a hero whose own power entailed her downfall.

Is Montgomery right that these three are not scientists, but shamans? In a revealing preface, Montgomery describes her own study of three young emus, Australian birds tall enough to look a woman in the eye. With minds so alien, she could not delude herself into even the partial empathy we feel for apes. Yet Montgomery realized that what

might be thought a respectable sum for a male scientist teaching on two continents, running a major research camp five hours' boat-ride into the rainforest, and incidentally mothering small children. Even Fossey, with her tortured warfare in defence of her gorillas, published a book and journal articles which will stand for years to come.

In her latest book, *Through a Window* (for review see *Nature* 348, 371; 1990), Goodall tells how she went to Gombe to recover from her second husband's death. In that time she came "intuitively . . . closer to the chimpanzees than ever before. For I was with them not to observe, to learn, but simply because I needed their company, undemanding and free of pity." That makes Montgomery's point: communion with nature goes far deeper and broader than the scientific attitude. It makes mine as well, that for 20 years before, and ten since, Goodall has gone notebook in hand, observing, thinking as well as feeling.

A different interpretation is that these three have led a paradigm shift in the study of animal behaviour. Goodall's insistence on the importance of the individual has now been accepted in primate studies, and resonates out into further and further fields. Kuhn showed us that such a paradigm shift is al-

ways resisted — but that it is an integral part of science, not a renunciation of science. Montgomery also quotes praise, concluding, "Jane's work has ushered in the glimmerings of a new way of doing science, a scientific outlook that draws upon the feminine emphasis upon individuality, relationships, and empathy".

Much is now written about women's science as different from men's. This may be so. But I suspect that an accurately attuned feeling for the organism will turn out to be not just female science, but good science. The important thing is not to lose either thought or feeling in an artificial dichotomy. In this present world, we humans threaten the survival of wild ecosystems and wild creatures including the great apes, our nearest relatives. We are

only just realizing how dependent even our own species is upon a fragile biosphere. We desperately need modern shamans: wise women and wise men. Our modern shamans, to be most effective, must also be scientists. If we insist on a dichotomy and champion either blind emotion or sterilized intellect, we risk deliberately letting go of our lifeline to survival — for the apes, or even for ourselves. □

Alison Jolly is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA.



Communion with nature — Hollywood's version of Dian Fossey's tragic life.

she felt for the birds was love — a special kind of love that has no basis in dependency, no expectation of return. It is this love, this communion she attributes to Goodall, Fossey and Galdikas.

Right. But then she spoils it by undervaluing their scientific contribution. She repeats disparaging (and unjustified) remarks about Goodall's Cambridge thesis. She only once mentions the 25-year monumental monograph *The Chimpanzees of Gombe* (Harvard University Press, 1986). She cites Galdikas as having only 40 publications, and that her one book so far is in Indonesian. This

On *homo loquens*

Paul Fletcher

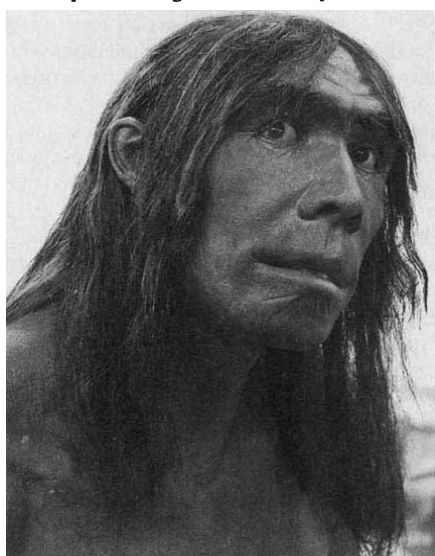
Uniquely Human: The Evolution of Speech, Thought and Selfless Behavior. By Philip Lieberman. *Harvard University Press*: 1991. Pp.210. \$27.95, £22.25.

CHIMPANZEES cannot talk. This indisputable fact is dealt with by linguists in different ways. The inability of species closely related to humans either to speak, or to learn a linguistic system of appropriate complexity, but expressed through some other medium (for example, manual signs) is for the (majority) chomskyan school a matter of some significance. It seems to confirm the distinct and unique status of humans and their language *vis-à-vis* our evolutionarily close relatives and their communication systems. Once a dichotomy between human language and animal communication is accepted, the research task for the chomskyan linguist (and, it has to be said, most others) is then the investigation of the principles of organization of human language in general and languages in particular. This research proceeds without reference to the evolution or present state of the anatomical structures or brain mechanisms which support these principles. Indeed, except in studies of language breakdown in adults, particularly adult aphasia, it is fair to say that brain-behaviour relationships are not of direct interest to many linguists.

There is however a menshevik tendency, of which Philip Lieberman has been a leading light, whose response to the uniquely human gift of tongues is to try to understand how and when the discontinuity between modern man and his predecessors came about. A major part of *Uniquely Human* reviews more than two decades of his research on this and related issues. Starting from the end-point — the current anatomy and physiology of the human speech-production apparatus — Lieberman calls into aid the comparative anatomy of the vocal mechanisms of extant primates, and interrogates the hominid fossil record. Crucial differences emerge between modern man, and his relatives and ancestors, in the shape of the supralaryngeal vocal tract and the position of the larynx. Because we are aware that chimpanzees do not talk, we can take it, or so the argument goes, that any archaic hominid whose laryngeal position and supralaryngeal vocal tract configuration resembles that of the modern chimpanzee could not talk either. And the inability to talk, in Lieberman's view, means the absence of those neural mechanisms that serve speech, and also the absence of syntax and the (related) neural mechanisms that serve it. So the nonhuman-like vocal tract of the La Chapelle-aux-Saints neanderthal, buried about 50,000 years ago in southwestern France, suggests that if he did have speech, it was not recog-

nizably human, and unlikely to be supported by the same brain mechanisms that support modern man's communication system.

This line of argument, attractive and intriguing as it may be, is not uncontroversial. Other researchers have probed Lieberman's chain of reasoning at various points. The conclusion that this particular neanderthal did not have speech rests on both a reconstruction of his vocal tract from basicranial measurements, and computer modelling of possible formant frequencies for vowels from this vocal tract. Both aspects of the investigation have been subject to detailed attacks, which in turn have provoked spirited ripostes. For a member of the linguistic majority, it is hard to see how the disputes can be resolved. In the absence of the critical data (which John Marshall once suggested, presumably not altogether seriously, would be



Commonwealth Institute

Could she speak? — life-size model of neanderthal female.

the discovery of a deep-frozen neanderthal who was susceptible to resuscitation) the comparative and evolutionary perspective insisted on by Lieberman would remain an interesting modern sidelight on speculations on the origin of language.

As it happens, this is not the end of the story. Lieberman's ideas on the representation, learning and breakdown of language in modern man, which are influenced by his evolutionary and comparative studies, are also provocative. For him, the availability of a human vocal tract guarantees important brain mechanisms. The brain has to be able to control the complex manoeuvres necessary for the control of the rapid articulatory movements of speech, which result in between 25 and 50 sounds being produced per second. In addition, the human perceptual apparatus has to be sophisticated enough to achieve resolution of the temporal and spectral characteristics of such an acoustic signal. So far, so good. The evolutionary advantage of a system that could voluntarily produce and perceive rapidly transmissible message is obvious. But linguists have long observed that there is a 'dual articulation' to the organ-

ization of speech: sounds are lawfully arranged in sequences, but then the words which consist of sequences of sounds are constructed in sentences which have a quite distinct set of rules (syntax) for their arrangement. Lieberman's assertion that there is nothing mysterious about syntax will come as a surprise to anyone familiar with the linguistic literature of the last 25 years. Syntax may be simple, in the sense that all languages have it, and all children learn it, but then from a similar perspective so is vision, or any other cognitive skill. For the organism the acquisition of syntax may be relatively effortless; the hard labour is understanding how it happens.

Of course, Lieberman's assertion is to some extent rhetorical. Although the story that links the anatomical characteristics of human-like vocal tracts to specialized brain mechanisms for the complexities of the motor control of speech production, and for the perception of rapidly transmitted sounds, is plausible, it is not at all clear how syntax might fit in to the account. He finesses the problem by proposing that the evolutionary basis for the brain mechanisms that underlie human syntactic ability is rapid vocal communication. So those mechanisms that had to develop and be selected to support speech "may have provided the preadaptive basis for rule-governed syntax". To buttress this argument he is almost required to play down the complexity of syntax, and ignore the quite distinct patterns of organization of sound and syntax in languages.

More substantive support for his view comes from studies of language-disordered individuals in whom verbal comprehension is very poor, but speech and syntax are accurate, though their language is described as semantically empty. This type of dissociation reported in sufferers from William's syndrome and some other conditions which also involve a more general and severe mental retardation, appears to suggest an uncoupling of speech and syntax from the rest of the linguistic and cognitive system. But there is contrary evidence, from research on both language-impaired children and adult aphasics, with normal intellectual abilities, which shows dissociations between speech and syntactic ability.

Uniquely Human is an uneven but rewarding and well-written book. The specialist will see it as swimming against the tide, and the nonspecialist should be warned that the author's perspective on *homo loquens* is not wholly representative of his field. The project is an ambitious one, and bound to have flaws. But in forcefully reminding those of his colleagues who may have forgotten, that there is a biological basis for what makes us human, Lieberman provides an alternative perspective which merits serious debate. □

Paul Fletcher is in the Department of Linguistic Science, The University of Reading, Whiteknights, PO Box 218, Reading, RG6 2AA, UK.

All in the mind

John C. Marshall

Cognitive Neuropsychology: A Clinical Introduction. By Rosaleen A. McCarthy and Elizabeth K. Warrington. *Academic*: 1990. Pp.428. Hbk £41, \$69; pbk £21, \$34.95.

Neuropsychological Impairments of Short-Term Memory. Edited by Giuseppe Vallar and Tim Shallice. *Cambridge University Press*: 1990. Pp. 524. £40, \$64.50.

A SIGNIFICANT part of current research in the 'decade of the brain' is characterized by an adverse ratio of hype to honest toil. All too often, the vogue for neural computation has brought forth neurophysiologically implausible models where the relationship between network performance and hard behavioural data ranges between false and mystical. There is excitement aplenty but rather little to show for the increased arousal levels. By contrast, much of 'normal' psychology is all too normal; yet another methodological twist to an experimental paradigm that produces small effects of limited relevance to anything other than the investigator's publication rate.

It is with some relief, then, that I turn to two new books on cognitive neuropsychology. Although I cannot claim to be an unbiased observer, it is nonetheless fair to remark that this discipline is intrinsically forced to confront a very practical reality: the existence of people with acquired brain damage whose cognitive impairment has seriously reduced their capacity for productive work and meaningful leisure. The standard neurological examination, combined with laboratory investigations and neuro-imaging techniques will (usually) reveal the proximal medical cause of the patient's condition. But it remains for the neuropsychologist to elucidate the patient's functional diagnosis, prognosis, and rehabilitation potential against a background of normal cognitive processing.

NEW JOURNALS ISSUE

This year, Nature's annual new journals review supplement will appear in the issue of 3 October. Publishers and learned societies are invited to submit journals for review, taking note of the following criteria:

- Journals that first appeared after June 1989 and issued at least four separate numbers by the end of April 1991 will be considered.
- The deadline for submission is end of May
- Journals covering any aspect of science are eligible, but those dealing with clinical medicine, engineering and pure mathematics are excluded, as are publications of abstracts
- Frequency of publication must be at least three times a year
- The main language used must be English.

When submitting journals for review, please send at least four different issues (the first, the most recent and any two others).

Cognitive Neuropsychology: A Clinical Introduction by R. McCarthy and E. Warrington is, quite simply, the clearest, most comprehensive, and up-to-date account of the discipline in print. The volume's coverage is exemplary for those areas where a substantial database exists: object and face recognition, spatial perception, voluntary action, perception, comprehension and production of language, reading and writing, arithmetic calculation, memory, and problem solving. As indicated, the book is organized by functional categories rather than disease entities; different aetiologies often provoke very similar behavioural impairments.

For each domain, McCarthy and Warrington first outline the basic empirical characteristics of how performance can break down within a particular subsystem. Much of the expository burden rests upon the detailed analysis of individual cases who show an especially 'pure' or extreme disorder (or small set of related symptoms). Group studies (or, better, series of well-investigated cases) are then used to reveal the (gross) loci of anatomical damage that has caused the problem. Each chapter concludes with a (brief) theoretical section in which deficits are related to a functional architecture of normal cognition; this architecture is usually expressed as a flow chart that diagrams how "different (processing) components are brought together to perform a specific task." Finally, functional pathology is interpreted as impairment to individual computational centres and the transfer of information between them.

Given the task that McCarthy and Warrington set themselves, it is difficult to imagine that it could have been better done. The book is, however, an introduction, as the title states. A few areas are ignored (disorders of time estimation and of music, for example). Cognitive aspects of personality and thought disorder after focal lesion receive short shrift; what little is included on these topics is hidden away behind attentional or problem-solving impairments. Prognosis and remediation are likewise not covered. There is good reason for these omissions: namely, a relative paucity of reliable and insightful studies. But I would have liked to see more attention paid to the evolution of cognitive deficits. Change in performance over time (manifest either as spontaneous recovery or as progressive deterioration) can provide dynamic constraints on theory that are difficult (or impossible) to obtain from any cross-sectional investigation.

Space limitations must also dictate selectivity in a book of this breadth; a point forcibly brought home by the fact that *Neuropsychological Impairments of Short-Term Memory*, edited by Giuseppe Vallar and Tim Shallice, can devote 500 or so pages to one neuropsychological symptom complex without padding or repetition. But repetition is the topic. Patients with left posterior parietal damage have severe difficulties in recalling

lists (random sequences of words, letters, or numbers, for example) immediately after hearing them. The phenomenon first assumed major theoretical significance when, in the 1960s, Warrington and Shallice documented a surprising dissociation: some patients with dramatically reduced auditory-verbal spans (tested by immediate repetition after one presentation) showed normal learning and long-term memory functions with multiple stimulus presentations.

This finding casts doubt upon what was then the standard account of short-term memory; that it was a necessary 'holding buffer', essential for the transfer of information into long-term storage. The controversy about the relationship between short- and long-term memory has rumbled along ever since, supplemented by an additional question: if the classical account is wrong, then what on earth is the natural function of a short-term, limited-capacity buffer? It is unlikely that the brain should contain a mechanism whose sole purpose is to repeat back telephone numbers once heard.

Vallar and Shallice have assembled a set of outstanding researchers who contribute chapters on just about every aspect of short-term memory and its neural correlates. In addition to patients with disproportionate difficulty on span tasks after focal lesion, the volume discusses the normal development of immediate memory in children, and the course of its deterioration in healthy older people. Much of the book contains new data and critical discussion of the most plausible hypothesis about the functional role of a short-term buffer: the buffer preserves a low-level phonological representation of sentences just in case the first pass interpretation is wrong (for example, "The teacher always taught by the new method passed the test"). Yet it remains extremely odd that some patients with very restricted spans perform well on tasks where one feels intuitively that a short-term verbatim record must have been preserved.

Such counterintuitive findings are, of course, one important reason why neuropsychological inquiry has played, and continues to play, such a dominant role in the development of cognitive psychology as a serious science. The extreme dissociations in performance seen after acquired brain damage are not random fracture lines. Rather, they reveal an underlying modularity of neuronal mechanisms that is often obscured by an apparent interaction of components in the brain when running smoothly. The painstaking experimental analyses reported in *Neuropsychological Impairments of Short-Term Memory* and in *Cognitive Neuropsychology: A Clinical Introduction* provide theoretical constraints that must be met by formal models of neural computation. □

John C. Marshall is at the Neuropsychology Unit, University Department of Clinical Neurology, The Radcliffe Infirmary, Oxford OX2 6HE, UK.

Avian arrangements

Jared M. Diamond

Phylogeny and Classification of Birds: A Study in Molecular Evolution. By Charles G. Sibley and Jon E. Ahlquist. Yale University Press: 1990. Pp. 976. Hbk \$100, £60.

Distribution and Taxonomy of Birds of the World. By Charles G. Sibley and Burt L. Monroe, Jr. Yale University Press: 1990. Pp. 1111. Hbk \$125, £75.

"I'm sick of hearing about that stuff. I no longer pay attention to anything those guys write."

"Why would anyone want to do something so boring as all that bird taxonomy?"

"Their results are the Revealed Truth, and you better believe it."

As these divergent opinions of three scientist friends of mine illustrate, taxonomy is a field noted for controversy. The opinions all refer to the same study, now summarized in two books, that used molecular methods to revolutionize taxonomy. These books make clear why taxonomy, though often despised as the most mundane science, inspired crucial insights in Darwin and continues to offer crucial new insights today.

As background, recall that the morphological criteria traditionally used in taxonomy are imperfect because they are subject to natural selection. Hence legacies of close relationship are often mimicked by adaptive convergence (for example, of whales and sharks) and obscured by adaptive divergence (for example, of New World vultures from storks). The same problems affect amino-acid sequences as taxonomic characters. DNA has the advantage that most DNA seems to be nearly neutral selectively and seems not to code for detectable proteins.

DNA-based taxonomic studies use two alternative methods: sequencing of short DNA segments, or single-copy DNA hybridization. The latter method involves measuring the difference between two entire genomes by the observed lowering of mixed melting point (about 1°C for a base-pair mismatch of one per cent). Theoretically, both methods should converge on the same answer with increasing length of the sequenced segment. This expectation is confirmed by multispecies comparisons of hominoid globin DNA sequences. But single-copy DNA hybridization has decisive practical advantages, because (in Roy Britten's words) it "is quicker than DNA sequencing and the results are more conclusive since the whole genome is averaged."

To date, by far the most extensive application of DNA hybridization to taxonomic problems has been the study of bird classification by Charles Sibley and Jon Ahlquist. From 1974 to 1986 they collected specimens from thousands of individual birds belonging to 1,700 of the world's 9,672 bird species, and they measured melting curves of 26,554 bird DNA hybrids. Birds have the advantage that, on the one hand, they have been intensively studied and almost all ex-

tant species have been described; and that, on the other hand, the severe aerodynamic constraints imposed by flight make traditional morphological criteria especially equivocal. This massive study is the basis of the two correspondingly massive books under review.

Phylogeny and Classification of Birds by Sibley and Ahlquist presents the immediate results of the study itself, and much more. Nearly half the text consists of excellent detailed accounts of general subjects that will interest many biologists who care not a whit about birds. These subjects include the organization of DNA, the (debated) selective neutrality of most mutations, practical details of DNA hybridization, problems in deducing relationships by any method, differing philosophies of taxonomy, the tempo of evolution, and the mutual insights derived from the knowledge of Earth history and of taxonomy. The rest of the book consists of the results for each group of birds, buttressed by detailed presentations of the data (nearly 400 sets of melting curves and phylogenetic trees).

Distribution and Taxonomy of Birds of the World by Sibley and Burt Monroe then goes through the world's 9,672 bird species, summarizing for each species its geographical distribution, preferred habitat, relationships, and classification based on the Sibley/Ahlquist conclusions. The classification is according to W. Hennig's cladistic principles: it reflects only the inferred branching pattern of phylogeny; taxonomic rank reflects only the inferred time of divergence; sister groups have coordinate rank; and degree of adaptive differences is not a consideration. The previous reference work most nearly similar in content, the 16-volume *Check-List of the Birds of the World* (Harvard Museum of Comparative Zoology, 1931–87) by James Peters and his successors, instead does consider degree of adaptive differences. Peters' set has the advantage of being far longer and more detailed because its unit is the subspecies, whereas Sibley's and Monroe's unit is the species. Peters' set has the disadvantage of having been written by many authors over the course of 56 years, resulting in somewhat varying standards and much now out-of-date material. Sibley's and Monroe's volume has the advantage of uniform standards, uniformly modern information, a large set of maps and gazetteer, and consistent evaluation for every species of its association in superspecies and its division into subspecies

groups. Every student of bird taxonomy will surely want to own both Peters and the new book by Sibley and Monroe, though students with limited budgets may have to content themselves with Sibley and Monroe alone.

Let us return to the criticism; "Why would anyone want to do something so boring as all that bird taxonomy?" Among the answers implicit in these volumes, here are three examples.

First, with true relationships established by non-morphological criteria, morphological evidence can now be used independently to study evolutionary convergence and divergence. The enormous richness of examples in these volumes will occupy evolutionary biologists for a long time. For instance, Sibley and Ahlquist discovered a marsupial-like radiation involving most Australian songbirds, a radiation that was previously unrecognized because those songbirds so closely resembled their unrelated Eurasian ecological equivalents, and because Australian songbirds are not united by any visible shared feature of morphology like the marsupial pouch. Storks and New World vultures prove to be closely related but strikingly divergent in adaptive characters, the former being convergent on herons, the latter on Old World vultures. Lest anyone fear that Sibley and Ahlquist have solved all problems of bird taxonomy, hundreds of major ones remain, involving relationships within families as well as between higher groupings. For instance, with more information the radiations of parrots and pigeons may come to tell tales as fascinating as the radiation of Australian songbirds.

Second, the effect (if any) of generation time on rates of evolution has been much debated. Sibley's and Ahlquist's results show clearly that demographic parameters have large effects on the rate of accumulation of selectively neutral mutations: the rate is higher in species with shorter generation times or with younger ages at first breeding. Because these two variables tend to be linked, it is still uncertain which variable is responsible. That issue, and the problem of the underlying mechanism, now become questions of great interest ripe for attack.

As a final example, all scientists face the dilemma of reconciling their desire for the truth with their personal egos and their instinctive defensive reactions to personal criticism. That dilemma is acute in cladistic taxonomy, a field notorious for its prevailing standard of vicious personal criticism, but the Sibley/Ahlquist study has attracted criticism notable for its viciousness even by the standards of taxonomists. It is thus all the more impressive that Sibley and Ahlquist in their book are conspicuously fair in presenting both the limitations and the advantages of their own and other results. In many places they cite and correct erroneous conclusions that they (as well as others) reached previously, and they trace how they became misled in each case. Their presentation can

serve as a model of how to handle criticisms and errors.

Who should buy which book? Anyone studying birds will want both books. They are long, crammed with information, and a bargain for the price. Anyone concerned with relationships within a group other than birds will at least want to study Sibley and Monroe, as a model of what taxonomy can achieve. The Sibley and Ahlquist book is for anyone in molecular evolution and evolutionary biology, for many other types of biologists, and for anyone (including historians) interested in how to do and present landmark science. □

Jared M. Diamond is at the University of California Medical School, Los Angeles, California 90024, USA.

Classical agriculture

Paul Halstead

The Ecology of the Ancient Greek World. By Robert Sallares. Duckworth/Cornell University Press: 1991. Pp.588. £42, \$84.

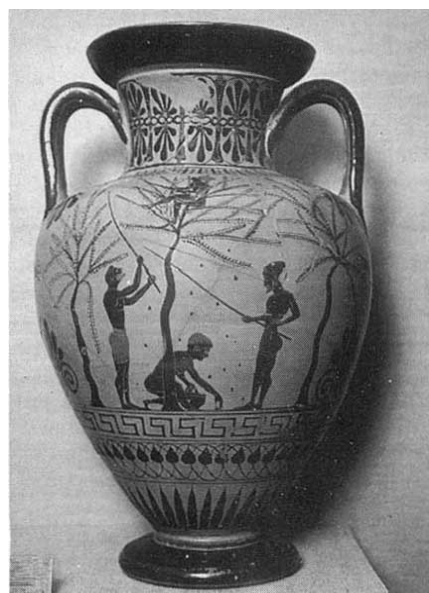
UNLIKE prehistorians, who readily pillage the natural sciences for insights in to the material record of the human past, ancient historians prefer the more familiar and superficially more tangible testimony of ancient texts. In *The Ecology of the Ancient Greek World*, Robert Sallares aims to cast some much-needed, multidisciplinary light on this dim recess of academia by writing "the first comprehensive ecological history of the ancient Greek world".

That Sallares transcends contemporary disciplinary parochialism is beyond dispute — detailed philological and historiographical argument competes for space with wide-ranging forays into archaeology, anthropology and evolutionary biology and ecology. With 800 footnotes and 2,000 bibliographic entries, *The Ecology of the Ancient Greek World* is unquestionably a serious work of scholarship, but never dull. Sallares has opinions on the irrationality of vegetarianism, the evolution of concealed ovulation in humans and the evolutionary prospects for AIDS. He explains the nature of causality and, in a scathing but witty assault on the traditional classical scholar, advocates comparative research as the historian's equivalent of the experimental method in science. He dismisses M. Finley, G. de Ste. Croix, M. Weber, K. Marx and F. Braudel on early Mediterranean towns, repeatedly admonishes 'New Archaeologists' for invasion-phobia, effectively demolishes E. Boserup's model of agricultural development and, at rather greater length, rewrites the evolutionary history of wheat. The historian is brought up to date on the extinction of dinosaurs, history of

bubonic plague and Dutch elm disease; ecologists should enjoy the pioneering work of Hegesandros on island biogeography.

Sallares defines ecology somewhat narrowly as the study of "the distribution and abundance of populations of living organisms in relation to their environment" and of "rates of flow of energy between different levels of food chains". Accordingly, the core of the book deals at length with the demographic and agricultural history of Greece during the first millennium BC. Recent evidence from intensive archaeological surveys, which has revealed drastic and widespread fluctuations in site numbers is taken to indicate equally drastic fluctuations in human population density. The problems of variable 'visibility' of different chronological horizons and of evident alternation between nucleated and dispersed settlement are acknowledged but then ignored — Sallares is, throughout, least convincing in his handling of archaeological evidence.

More particularly, he follows A. Snodgrass in arguing that increasing numbers of burials during the early first millennium BC mark a steep rise in population; he perceptively criticizes but fails to refute T. Morris' suggestion that this trend reflects archaeologically visible burial of an increasing proportion of the population. A lengthy digression on skeletal evidence for the age and sex structure of the population is far too inconclusive to clarify changes in population size. More intriguing is the proposition that early Greek society in the first millennium BC was characterized by a system of age-class organ-



Ronald Sheridan

Images of nature — vase depicting olive harvester from the fourth century BC.

ization, which served to restrain population growth by delaying marriage for males (delayed marriage for females would have been more effective). Sallares seems to demolish his own case, however, by arguing that in Attica the age-class system was falling into disuse at a time when population growth was

slowing down.

On agriculture, Sallares makes perceptive comments on the ecological and economic context of recent farming practices, and on the dangers of uncritical extrapolation from present to past. Ultimately, however, his arguments on the nature and scale of classical land use are sloppy, failing to draw a clear distinction between what was ecologically or technologically possible, what would have been economically viable or 'rational' and what may be implied by scarce historical and archaeological evidence. The unorthodox claim that classical Athens was basically self-sufficient in staple grains is probably correct, but hardly substantiated. Even more unjustified is the assertion that the steep rise in human population during the first millennium BC was made possible by significant improvements in agriculture. Sallares justly criticizes the traditional emphasis on improvements in agricultural technology, but his claim that the displacement of glume wheats by free-threshing wheats reflects the improved productivity of the latter is quite unsubstantiated. Even if classical archaeologists bothered to collect charred crop remains, grain size is affected by interannual and local variation in growing conditions, by crop-cleaning methods and by charring regime, and is only indirectly related to area yield. Given that naked wheats had remained in cultivation on a small scale for millennia throughout Europe, their ultimate displacement of the more storable but troublesome glume wheats surely reflects a shift of emphasis in risk-buffering behaviour from local self-sufficiency and storage to regional integration and exchange. Furthermore, the assumption that classical civilization was underpinned by surplus resulting from higher-yielding crops reflects the same misplaced faith in progress which Sallares effectively condemns in earlier works. In fact, in an area of high agricultural risk such as southern Greece, a 'normal surplus' will have been an inevitable by-product of any viable farming economy and an investigation of changes in access to land, labour and produce would probably have been far more illuminating.

Despite these criticisms, *The Ecology of the Ancient Greek World* is informative and thought-provoking. Sallares' inability to leave any intellectual stone unturned does nothing for the clarity of his argument, but it also makes him eminently readable and his evident enthusiasm for history and ecology is very engaging. This unusual book, with its blend of references to scientific papers from the first millennium BC and 1980s AD, will interest a wide readership. For ancient historians, to whom it is primarily addressed, it should be compulsory reading and those not yet beyond redemption will learn much from the experience. □

Paul Halstead is in the Department of Archaeology and Prehistory, Sheffield University, Sheffield S10 2TN, UK.

Flaming fronts of fire

J. A. J. Gowlett

Burning Bush: A Fire History of Australia. By Stephen J. Pyne. *Holt*: 1991. Pp.520. £19.95, \$27.95.

FIRE strikes deep chords in the human psyche. Terror of wildfire is balanced by love of the essential hearth — the heart of figurative speech in all language. In one of many Australian aboriginal fire myths, giant kangaroos attack humans until they are driven off by fire, which the people then keep safely in the bole of a eucalyptus tree. Fire history encompasses much that is comparably odd or romantic, so it becomes a bold undertaking to weld together an account embracing both natural phenomena and human influences on a continental scale. Stephen J. Pyne had the enthusiasm to do this for the Americas, and now has turned his attention to Australia.

Throughout *Burning Bush: A Fire History of Australia*, Pyne deals in colourfully fiery language. After the break-up of Gondwana, Australia followed a separate destiny, drifting through the oceans on its own evolutionary path. Fire had touched other continents, but here, Pyne explains, it was something much more deep-rooted: fire branded Australia, "pervasive, persistent, singular". A constant and essential theme of the book is that those species which have shaped Australian fire have been, in turn, shaped by it. First among all these is the eucalyptus, the gum, in all its variety of more than 600 species. Next perhaps is the acacia; then humans — the aborigines, the early white settlers, and finally the farmers, foresters and government commissions.

Such a theme demands almost an epic, and in spite of a disclaimer, Pyne provides it. At times he writes a veritable fire poetry. As in Homer, key phrases become familiar, favourite stories recur; and also, there are subsidiary plots, followed one by one to their logical conclusions before the main theme is rejoined. The volume is divided into books; fire sweeps, spouts and glows on every page. Here there is a potted fire history of Britain, because of the British colonization; there, of California because of its similar Mediterranean environment; of the outlying islands because they were not shaped by human occupation.

This concentrated stream of fire is aimed not at one audience, but in the author's own desire two, those of America and Australia. I would go further, and say that the book is suitably pitched for both general and academic audiences anywhere in the English-speaking world, provided only that they have the stamina for such a *tour de force*. The vigorous metaphorical style is not, however, a cover for a work lacking in content. The

author covers his ground in remarkable detail, backing up his text with many pages of bibliographic notes. He starts by working systematically through the natural environment, emphasizing the variety of its fire potential. In heath, for example, more than 90 per cent of vegetation is available for burning, but in *Acacia brigalow* less than 0.1 per cent. Then there is the coming of the aborigines, the 'firestick farmers' who burned regularly in the course of everyday life, flushing game, stimulating new growth and undoubtedly altering the fire regimes dramatically. Descriptions of aboriginal fire practices and beliefs are treated well. After early European settlement come the later phases or organized fire control and legislation.

For those concerned with the past, the most interesting question of all is likely to be the interaction of climatic change, vegetation change, the coming of the aborigines, and the disappearance of the megafauna. The sclero-

lives. These follow the past parade of Red Tuesday, Black Thursdays, Fridays, and Sundays. No management strategy is feasible that will expunge fire entirely. Pyne draws a contrast with the approaches taken in North America, which concentrate more on suppression. Time and again Australians have reverted to the notion that regular controlled burning is necessary to prevent inordinate amounts of fuel building up. Hence the dictum of the forester Harry Luke, that fire control means fuel control.

A book such as this could easily give the impression that fire is everything, the dominant influence of all; but this is not quite what the author says, in spite of his devotion to his subject. If occasionally it looks as if we are getting just one side of a story, Pyne always comes back with a critical analysis. Fire is not paramount in Australia. There are areas where rainforest has reintroduced itself since the last glaciation. Some types of vegetation



Fanning the flame — an aborigine kindles dry grass in the ancient way.

morphic eucalyptus has expanded greatly during the last glaciation, much reducing the cover of rainforest. The first human occupation goes back over 40,000 years; some of the megafauna survived much after then. Was climate the main agent of change, or did the more regular and systematic burning of the 'firestick farmers', coupled with their hunting, contribute to the disappearance of the giant kangaroos and wombats?

Questions of the present are not much easier to answer: the problems of fire are not past history. During the last major fires on Ash Wednesday 1983 fire tornadoes were recorded, one ranging up to 375 metres high; 180 fires started in the southeast on the one day, ten raging out of control, burning more than 350,000 hectares and claiming over 70

have evolved so that fire cannot flourish within them. Victoria, which has suffered more than half the total economic damage from bushfires, represents less than three per cent of Australia's landmass. Modern human influence has made more of fire, made its escape from controlled fire to bushfire most costly. Objectively the true significance of fire is difficult to measure. For Professor Pyne, and for the aborigines in their 'dreamtime', its importance can hardly be exaggerated. In the face of a bushfire, who would venture to say that they are wrong? □

J. A. J. Gowlett is in the Department of Archaeology, University of Liverpool, The Hartley Building, P.O. Box 147, Liverpool L69 3BX, UK.

Australian Overseas Information Service

Le style c'est l'homme

Walter Gratzer

Contrasts in Scientific Style: Research Groups in the Chemical and Biochemical Sciences. By Joseph S. Fruton. *American Philosophical Society*: 1990. Pp.473. \$40.

HISTORIANS are apt to justify their calling by recourse to the old saw that those who do not learn the lessons of history are fated to relive them. Joseph Fruton's account of the tensions and conflicts that lay beneath the heady triumphs of chemistry during its golden heyday in Wilhelmine Germany invites inevitably the comparison with the rise of molecular biology in our own time. The same faint odour of venality spread through the universities, for instance, as the prospects of commercial exploitation drew nearer; the same surly resentments sprang up in the breasts of research students, serving a tyrannical master in the first organized research teams.

There are also undeniably differences. Even the most olympian of German professors plied his trade at the bench — the first among unequals. Robert Bunsen, one of the giants of his (or any) time, according to a contemporary quoted by Fruton, "occupied himself very actively with beginners. He showed them each manipulation and everything involved in an analysis." And here is W. H. Perkin's recollection of Adolf von Baeyer in action: "It was [my] good fortune to be allowed to take part and watch these experiments, and the experience of the experimental skills with which Baeyer carried out these very ticklish operations [the synthesis of a series of explosive acetylene compounds], with no other apparatus than test-tubes and glass rods, is to this day a very lively recollection."

Fruton's purpose is to explore how the relations that the head of a research group establishes with the workers in his charge relates to the success of the enterprise. The first of the titans, who deployed a small army of workers with its rigorously defined chain of command to overwhelm the chosen sector of his science, was Liebig. His laboratory at Giessen prefigured the type of operation now commonplace at Caltech and MIT, except that supervision was closer than Liebig's modern inheritors would regard as feasible or today's students and postdocs as tolerable. Liebig harried his students ceaselessly; Fruton quotes a passage from his memoirs, which concludes: "We worked from day-break until nightfall, distractions and diversions were not available in Giessen." Even his practice — unusual at that time — of sometimes allowing his students to publish work under their own names in his house journal, the *Annalen*, was calculated: "The

people who work with me", he confided to one of his associates, "publish under their own names, even if I have helped them. If it is something good, a part of the credit is ascribed to me and I do not need to defend the mistakes. You understand?" It was indeed arguably more important to have been a *Liebig Schüler* than to have done work of substance.

By contrast, Hoppe-Seyler, intolerant and contentious as he famously was, showed exemplary generosity and loyalty towards his associates. Hofmeister was another whose patronage was warm and disinterested. Baeyer, who built up the most influential research school of his time, emerges as an awesome, yet inspiring figure, one whose approval was most eagerly sought. Legends sprang up about him; Fruton eschews the more frivolous anecdotes, but it seems a pity that he denies his readers the vision of the Old Man's delight when the crystals of the first terpene appeared in the tube, and the famous bowler, which left his head only in rare moments of high emotion, was raised twice in salute.

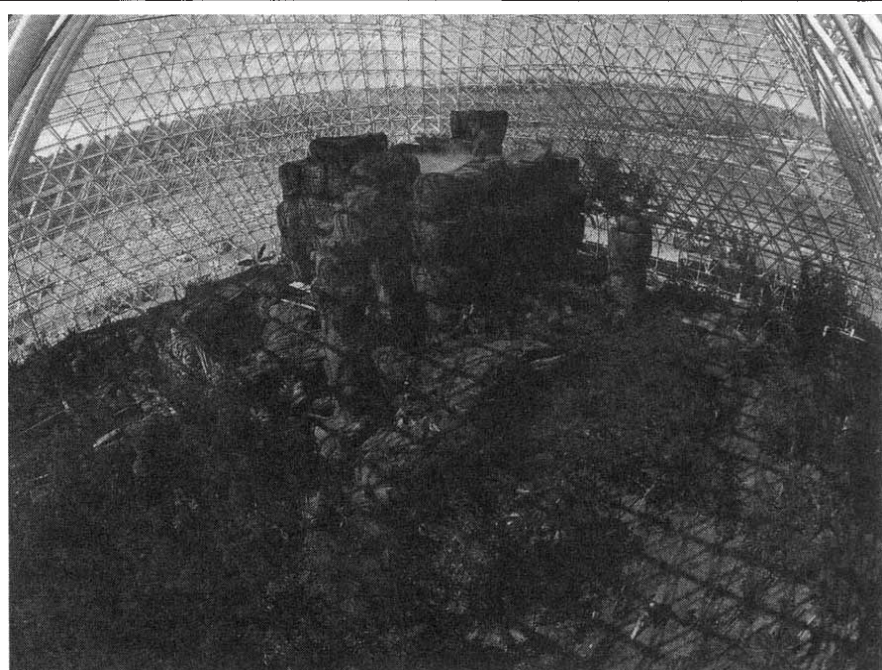
Baeyer's greatest and unswervingly devoted student was Emil Fischer, an austere and unattractive personage and a despot when he acceded to his own research group. Fruton believes that his unforthcoming character may have owed much to the ravages of chronic chemical poisoning, for he had exposed himself continuously and in an ill-ventilated laboratory, to phenylhydrazine and to volatile mercury compounds. There is also here a description of Fischer's first effort as a graduate student, which must surely have left a spiritual scar: he was required to effect the reduction of half a kilogram of mel-

litic acid with sodium amalgam — made from no less than 25 kilograms of mercury. He was carrying the flask that contained his hard-won product when the floor of the laboratory gave way and the whole lot was lost.

Fruton traces the complex and uneasy relation between organic chemistry and biochemistry, or physiological chemistry (*Schmierchemie* to the purists) as it was then called, and the kinds of cul-de-sac in which so many of the early efforts ended. "How little of what I learned with great pains", Otto von Fürth wrote poignantly, "has importance today, and how many of the problems that the previous generation of physiologists fought about so passionately have lost for us any sense or significance."

The last chapter of Fruton's magisterial work brings the story, by way of Gowland Hopkins, Warburg and Krebs to the biochemistry and molecular biology of the present day. It ends with 140 pages of appendices, in which are recorded the names of all those (so far as Fruton's excavations of the archives are able to determine them) who toiled at some time in the seven great laboratories that serve as his exemplars. Here are gathered the celebrated and the obscure — men (no women) who became professors, industrial chemists, shopkeepers and landowners — here one who fell at 30 in the Great War, there an octogenarian, who perished in Theresienstadt. It is a *memento mori* and a monument to the foot-soldiers of science; let us hope that one day a successor of Joseph Fruton's will do the same for us. □

Walter Gratzer is in the MRC Muscle and Cell Motility Unit, King's College, Drury Lane, London WC2B 5RL, UK.



Small-scale rainforest — Biosphere 2, a complete self-sustaining ecosystem modelled on Biosphere 1 (the Earth), has been constructed in Arizona and will be home to eight scientists for two years. *Biosphere 2: The Human Experiment* by John Allen is published by Penguin at \$16.95, £9.99.